

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comogen Ltd.

OW protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 59 Seconds

[without alignments]  
1431.893 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1639

Sequence: 1 MESTVRLTTLAALPLVNS.....TTCVSGTCCXNDYSGQL 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Genesegp29Jan04:\*  
2: genesegp1980s:\*  
3: genesegp2000s:\*  
4: genesegp2001s:\*  
5: genesegp2002s:\*  
6: genesegp2003as:\*  
7: genesegp2003bs:\*  
8: genesegp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1659	100.0	299	2	AAW04928
2	1659	100.0	299	2	AAW63624
3	1178	71.0	305	2	AAW44854
4	1178	71.0	305	2	AAW41929
5	1159	69.9	305	2	AAW15271
6	1159	69.9	305	2	AAW25428
7	1159	69.9	305	2	AAW28295
8	1159	69.9	305	2	AAW25525
9	1159	69.9	305	2	AAW25464
10	1159	69.9	305	2	AAW27963
11	1159	69.9	305	2	AAW42063
12	1159	69.9	305	2	AAW37150
13	1159	69.9	305	2	AAW67388
14	1159	69.9	305	2	AAW01502
15	1159	69.9	305	2	AAW388471
16	1159	69.9	305	2	AAW44266
17	1159	69.9	305	2	AAW46616
18	1159	69.9	305	2	AAW03660
19	1159	69.9	305	4	ABW00014
20	1146	69.1	305	2	AAW78352
21	1146	69.0	305	2	AAW78356
22	1143	68.9	305	2	AAW78359
23	1142.5	68.8	306	2	AAW44269
24	1142	68.8	305	2	AAW28300
25	1142	68.8	305	2	AAW78355

26	1142	68.8	305	2	AAW78358	AAW78358 Humicola
27	1138	68.6	305	2	AAW78353	AAW78353 Humicola
28	1137	68.5	305	2	AAW78360	AAW78360 Humicola
29	1137	68.5	305	2	AAW78354	AAW78354 Humicola
30	1137	68.5	305	2	AAW78357	AAW78357 Humicola
31	1134	68.4	305	2	AAW28818	AAW28818 H. insol
32	1127	67.9	286	2	AAW57420	AAW57420 Humicola
33	1122	67.6	200	2	AAW53968	AAW53968 Thielavia
34	1120	67.5	284	2	AAW04137	AAW04137 Mutant 43
35	1119	67.5	284	2	AAW04129	AAW04129 Mutant 43
36	1117	67.3	200	2	AAW53979	AAW53979 Thielavia
37	1114	67.1	284	2	AAW04141	AAW04141 Mutant 43
38	1113	67.0	284	2	AAW04140	AAW04140 Mutant 43
39	1112	67.0	284	2	AAW04130	AAW04130 Mutant 43
40	1112	67.0	284	2	AAW04128	AAW04128 Humicola
41	1110	66.9	305	2	AAW78361	AAW78361 Humicola
42	1108	66.8	284	2	AAW04131	AAW04131 Mutant 43
43	1108	66.8	284	2	AAW04135	AAW04135 Mutant 43
44	1107.5	66.8	306	2	AAW44270	AAW44270 Hybrid DN
45	1107	66.7	284	2	AAW04132	AAW04132 Mutant 43

#### ALIGNMENTS

RESULT 1

ID AAW04928 standard; protein; 299 AA.

AC AAW04928;

DT 20-MAY-1997 (first entry)

DE Cellulytic enzyme #4 of the invention.

XX Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;

XX plant cellulase; catalytic region; textile; backstaining; bio-polishing;

XX stone-washing; cellulosic fabric; colour clarification; debarkation;

XX cell wall degradation; paper pulp; debarking; fibre modification;

XX enzymatic de-inking; drainage improvement.

XX Thielavia terrestris.

XX M09629397-AL.

XX 26-SEP-1996.

XX 18-MAR-1996; 96WO-DK00105.

XX 17-MAR-1995; 95DK-00000272.

XX 08-AUG-1995; 95DK-00000885.

XX 08-AUG-1995; 95DK-00000886.

XX 08-AUG-1995; 95DK-00000887.

XX 08-AUG-1995; 95DK-00000888.

XX 12-FEB-1996; 96DK-00000137.

XX (NOVO) NOVO-NORDISK AS.

XX Schuelein M, Andersen LN, Lasser SF, Kaupinen MS, Lange L;

XX Nielsen R, Ihara M, Takagi S;

XX WPI: 1996-443173/44.

XX N-PSDB; AAT39050.

XX New endo-glucanase enzyme preparations - contg. conserved catalytic

XX regions, useful for treating fabrics, textiles, plant material or paper

XX pulp.

XX Claim 72, Page 161-163; 316pp; English.

XX AAW04925-M04932 represent the enzymes of the invention. These enzymes

XX possess cellulytic (particularly endoglucanase) activity. Cellulytic

XX enzymes are involved in the the hydrolysis of cellulose, and are

CC synthesised by a large number of microorganisms and plants. The enzymes  
CC of the invention containing the conserved catalytic regions (such as  
CC AAM04913) exhibit improved performance, e.g. 50 times higher performance,  
CC compared to multiple domain enzymes. The enzymes can be used for the  
CC treatment of fabrics or textiles, preferably for preventing backstaining,  
CC for bio-polishing or for stone-washing cellulosic fabric. They can also  
CC be used to provide colour clarification for laundry. The enzymes can also  
CC be used for the degradation or modification of plant material, such as  
CC cell walls. They can also be used in the treatment of paper pulp  
CC preferably for debarking, defibration, fibre modification, enzymatic de-  
CC laking or drainage improvement

SQ Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 8,6e-119;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60  
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60  
QY 61 LSDPNVSGCGSGSAVSCADOTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120  
DB 61 LSDPNVSGCGSGSAVSCADOTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120  
QY 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGSLPGAQYGGISSRDQDS 180  
DB 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGSLPGAQYGGISSRDQDS 180  
QY 181 FPAPLKPQCQKRFMFQNDNPTFTFOVQCRAIIVASGCKRNDSSFVPTPPSGANG 240  
DB 181 FPAPLKPQCQKRFMFQNDNPTFTFOVQCRAIIVASGCKRNDSSFVPTPPSGANG 240  
QY 241 GTGPTSTAPSGSGTSPGSGGCTSQKMAQCGIGFSGCTTCVSGTTQKLNIDYISQCL 299  
DB 241 GTGPTSTAPSGSGTSPGSGGCTSQKMAQCGIGFSGCTTCVSGTTQKLNIDYISQCL 299

## RESULT 2

AAM63624 standard; protein; 299 AA.

XX AAM63624;

XX 21-SEP-1998 (first entry)

XX Monocomponent endoglucanase enzyme.

XX Monocomponent endoglucanase; cellulolytic enzyme; garment; abrasion;

XX abraded looking jeans; fungus; Thielavia terrestris; stone-washed.

XX Thielavia terrestris.

XX EP843041-A1.

XX 20-MAY-1998.

XX 30-MAY-1997; 37EP-00610021.

XX 13-NOV-1996; 96DK-00001276.

XX (NOVO ) NOVO-NORDISK AS.

XX Lund H, Kalum Jr

XX WPI; 1998-263308/24.

XX N-PSDB; AAV39C96.

XX Pair of jeans - with abraded look.

XX Claim 16; Page 10-12; 16pp; English.

CC This represents a monocomponent endoglucanase. This is a cellulolytic  
CC enzyme belonging to the fungal species Thielavia terrestris. This can be  
CC used in the process of invention of providing a pair of jeans made from  
CC dyed twill fabric that has localised variations in the colour density of  
CC the fabric. The process involves providing the jeans with a stone-washed  
CC or abraded look, where the variation corresponds to a delta remission  
CC value (Delta R) higher than 11, and the reflection of a first area of the  
CC jeans fabric is less than 12 percent. The process comprises selecting the  
CC desired textile cutting pattern for jeans garment, positioning the  
CC pattern onto newly manufactured dyed twill fabric, cutting the jeans  
CC garment parts, sewing the pair of jeans, optionally subjecting the pair  
CC of jeans to a desizing treatment, and subjecting the pair of jeans to an  
CC abrasion treatment with an efficient amount of a cellulolytic enzyme in  
CC an aqueous medium essentially free of bleaching chemicals. The abraded  
CC look near stitching is different from the abraded look far from stitching

SQ Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 8,6e-119;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60  
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYMDCKPCSCAMPKAAVSQPVYACDANFOR 60  
QY 61 LSDPNVSGCGSGSAVSCADOTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120  
DB 61 LSDPNVSGCGSGSAVSCADOTPMAYNDNLAYGPAATSIAGSESSWCACALYLTSSGP 120  
QY 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGSLPGAQYGGISSRDQDS 180  
DB 121 VAGKTWVVGSTSTGDLGSGNOFDIAMPGGVGIFNGCSSQFGSLPGAQYGGISSRDQDS 180  
QY 181 FPAPLKPQCQKRFMFQNDNPTFTFOVQCRAIIVASGCKRNDSSFVPTPPSGANG 240  
DB 181 FPAPLKPQCQKRFMFQNDNPTFTFOVQCRAIIVASGCKRNDSSFVPTPPSGANG 240  
QY 241 GTGPTSTAPSGSGTSPGSGGCTSQKMAQCGIGFSGCTTCVSGTTQKLNIDYISQCL 299  
DB 241 GTGPTSTAPSGSGTSPGSGGCTSQKMAQCGIGFSGCTTCVSGTTQKLNIDYISQCL 299

## RESULT 3

AAM44854 standard; protein; 305 AA.

XX AAM44854;

XX 31-JUL-1998 (first entry)

XX Humicola insolens cellulase NCB4 protein.

XX Humicola insolens; NCB1; NCB2; NCB4; cellulase; expression vector;

XX promoter; signal sequence; terminator; amylase; lipase; protease;

XX phytase.

XX Humicola insolens.

XX key 1, 21 Location/Qualifiers

XX Peptide /label= signal

XX Protein 22, 305 /label= Cellulase\_NCB4

XX W09803667-A1.

XX 29-JAN-1998.

XX 24-JUL-1997; 97WO-JP002560.

XX 24-JUL-1996; 96UP-00195070.



LT 25-MAR-2003 (revised)  
 LT 18-FEB-1992 (first entry)  
 DE Humicola insolens DSM 1800 endoglucanase.  
 XX Humicola insolens.  
 KW cellulase; cellulose.  
 OS Humicola insolens.  
 FH Key  
 FT Peptide Location/Qualifiers  
 FT 1..21  
 FT /label= signal  
 FT Protein 22..305  
 FT /label= endoglucanase  
 XX MO9117243-A.  
 XX PD 14-NOV-1991.  
 XX PF 09-MAY-1990; 90DK-00001159.  
 XX PR 09-MAY-1990; 90DK-00001159.  
 XX PR 22-APR-1991; 91DK-00000736.  
 XX PA (NOVO) NOVO-NORDISK AS.  
 XX PI Rasmussen G, Mikkelsen JM, Schulten M, Packar SA, Hagen F,  
 PI Hjort C, Hestrup S;  
 XX WPI; 1991-353765/48.  
 DR N-PSDB; AAQ14856.  
 XX Cellulase prep. comprising endoglucanase enzyme - used in detergents for  
 PT cellulose-contg. fabrics or to improve drainage of paper pulp.  
 XX Claim 6; Page 48; 67pp; English.  
 CC The cellulase isolated from H. insolens has mol.wt. ca. 43kD by SDS-PAGE  
 CC and an isoelectric point of ca. 5.1. The enzyme can be used in detergents  
 CC for softening, soil removal and colour clarification. It can reduce the  
 CC harshness of cellulose-contg. fabrics. (Updated on 25-MAR-2003 to correct  
 CC PA field.)  
 XX SQ Sequence 305 AA;  
 Query Match 69.9%; Score 1159; DB 2; Length 305;  
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;  
 QY 1 MRSTPVLRTTLAALPLVNSAASGSGSTRYWDCKKPCSCAPKGAAYGQPYACDANFOR 60  
 DB 1 MRSSPLPSAVVAALPVIALAA--DGRSTRYWDCKKPCSCAPKGAAYGQPYACDANFOR 58  
 QY 61 LSDEFVNSGNC-GGSAVSCADOTPMAYNDNLAYGFAATSIAGSSESSWCCACALTFPSG 119  
 DB 59 ITDPDAKSGCEPGVAVSCADOTPMAYNDNLAYGFAATSIAGSSESSWCCACALTFPSG 118  
 QY 120 PVAKMTVVOSTSTGDLGSLNDFIAPGGGVIIFNGCSQFGLPAGVQYGISRDQCD 179  
 DB 119 PVAGKMTVVOSTSTGDLGSLNDFIAPGGGVIIFNGCSQFGLPAGVQYGISRDQCD 178  
 QY 180 SFPAPLPGCKMRFQNDNPTFFQVQCPALIVARSGCKRNDSSFP-VTPPSGG 238  
 DB 179 RFPDAPLPGCKMRFQNDNPTFFQVQCPALIVARSGCKRNDSSFP-VTPPSGG 238  
 QY 239 NGGTGPTSTAPSGSOTS-----PGGSGCTSQKMAQCGIGSGCTTCVSGTTCOKLN 292  
 DB 239 SSVPNQPTSTSTSTSTSSPVPQPTTSGCTAERMAQCGIGSGCTTCVSGTTCOKLN 298  
 QY 293 DYSQCT 299  
 DB 299 DMYQCT 305

RESULT 6  
 AAR25428  
 ID AAR25428 standard; protein; 305 AA.  
 XX AAR25428;  
 AC AAR25428;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 15-JUN-1993 (first entry)  
 DE Cellulase contained in a detergent compn.  
 XX Endoglucanase; immunoreactive; Humicola insolens; cleaning; softening.  
 XX Humicola insolens.  
 OS Humicola insolens.  
 XX EP495554-A1.  
 XX PD 22-JUL-1992.  
 XX PF 15-JAN-1992; 92EP-00200101.  
 XX PR 16-JAN-1991; 91EP-00870006.  
 XX PR 06-NOV-1991; 91EP-00202881.  
 XX PA (PROCTER & GAMBLE CO.  
 XX PI Conventa AC, Busch A, Baack AC;  
 XX WPI; 1992-243405/30.  
 DR P-PSDB; AAR25428.  
 XX Detergent compn. for improving cleaning and performance - is composed of  
 PT quat. ammonium cpd. and high activity cellulose e.g. homogeneous  
 PT endoglucanase.  
 XX Claim 10; Page 19; 28pp; English.  
 CC The cellulase was used in a detergent compn. contg. a quaternary  
 CC ammonium cpd. The compn. gives a detergent with good cleaning and  
 CC softening performance due to the synergistic effect of the ammonium and  
 CC cellulase components. The cellulase is a homogeneous endoglucanase  
 CC component which is immunoreactive with an antibody raised against a  
 CC cellulase purified from Humicola insolens DSM 1800. See also AAR25429.  
 CC (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to  
 CC correct OS field.)  
 XX SQ Sequence 305 AA;  
 Query Match 69.9%; Score 1159; DB 2; Length 305;  
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;  
 QY 1 MRSTPVLRTTLAALPLVNSAASGSGSTRYWDCKKPCSCAPKGAAYGQPYACDANFOR 60  
 DB 1 MRSSPLPSAVVAALPVIALAA--DGRSTRYWDCKKPCSCAPKGAAYGQPYACDANFOR 58  
 QY 61 LSDEFVNSGNC-GGSAVSCADOTPMAYNDNLAYGFAATSIAGSSESSWCCACALTFPSG 119  
 DB 59 ITDPDAKSGCEPGVAVSCADOTPMAYNDNLAYGFAATSIAGSSESSWCCACALTFPSG 118  
 QY 120 PVAKMTVVOSTSTGDLGSLNDFIAPGGGVIIFNGCSQFGLPAGVQYGISRDQCD 179  
 DB 119 PVAGKMTVVOSTSTGDLGSLNDFIAPGGGVIIFNGCSQFGLPAGVQYGISRDQCD 178  
 QY 180 SFPAPLPGCKMRFQNDNPTFFQVQCPALIVARSGCKRNDSSFP-VTPPSGG 238  
 DB 179 RFPDAPLPGCKMRFQNDNPTFFQVQCPALIVARSGCKRNDSSFP-VTPPSGG 238  
 QY 239 NGGTGPTSTAPSGSOTS-----PGGSGCTSQKMAQCGIGSGCTTCVSGTTCOKLN 292  
 DB 239 SSVPNQPTSTSTSTSTSSPVPQPTTSGCTAERMAQCGIGSGCTTCVSGTTCOKLN 298

QY 293 DYSOCL 299  
|:|  
Db 299 DWHQCL 305

RESULT 7  
AAR28295  
ID AAR28295 standard; protein; 305 AA.

XX AAR28295;  
AC  
XX 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 02-APR-1993 (first entry)

DE Sequence of ~ 43 kD endoglucanase.

XX Detergent additive; protease; enzyme; endoglucanase.

XX Humicola insolens; DSM 1800.

XX Key Location/Qualifiers

FT Peptide 1..21  
FT /label= signal

FT Protein 22..305

PN M09218599-A1.

XX 29-OCT-1992.

XX 10-APR-1992; 92MO-DK000116.

XX 22-APR-1991; 91DK-00000737.

XX (NOVO) NOVO-NORDISK AS.

XX Wagner P. Tsuchiya R;

XX WPI; 1992-382092/46.

XX N-PSDB; AAQ30067.

XX Detergent additive concy. cellulase and specific protease - which does not degrade the cellulase during storage and clarifies the colour of dyed cellulosic materials.

XX Disclosure; Page 8-9; 15pp; English.

XX The ~43 kD endoglucanase derived from Humicola insolens, DSM 1800 is described in detail in co-pending Danish patent application No. DK 1159/90. As a detergent additive, it has a higher degree of specificity than Bacillus lentus serine protease. The term "higher degree of specificity" is defined as a protease which conditions degrades human insulin to fewer components. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1,4e-80; Indels 10; Gaps 4;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSPVYACDANFOR 60

Db 1 MRSSPLPSPAVVALLPVIALAA--DGRSTRWDCCRRSCAMKAPVAVQPFSCANFOR 58

QY 61 ISDPNVQSGCN-GGSAYSCADQTPMAVNDNAYGFAATSIAGSSSSNCCACALFTFSG 119

Db 59 ITDPAKSGCERGGVAVASCADQTPMAVNDNAYGFAATSIAGSSSSNCCACALFTFSG 118

QY 120 PVAGKTVYVGTSTGSLGSGNOFTIAMRGSGVGRVNCSSQSGGIPKAGVIGISSNDCCD 179

Db 119 PVAGKTVYVGTSTGSLGSGNHEDINIGGGVGIPIQCTPGFGLPGQRVGGISSNRECCD 178

QY 180 SFPAELKPGCWRPFDMFQADNPFTETFOOVCCPAETVARSGCKRNDSSFP-VFTPPSGG 238  
|:|  
Db 179 RFPDLKRGCVWRPFDMFQADNPFSFRVQCPALVARTGCRNDGNFPAVQIPSSST 238

QY 239 NGGTGPTSTAPSGQTS-----PGGSGCTSQKMAQCGGIGFSGCTTCVSGTTCQLYN 292  
|:|  
Db 239 SSPVNGPTSTSTSTSTSTSPVQPTTPSGCTAERMAQCGGNGMSGCTTCVAGSTCTKIN 298

QY 293 DYSOCL 299  
|:|  
Db 299 DWHQCL 305

RESULT 8  
AAR25525  
ID AAR25525 standard; protein; 305 AA.

XX AAR25525;

XX 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-JAN-1993 (first entry)

XX Humicola insolens DSM 1800 endoglucanase.

XX Detergent; washing powder; cellulase; softening clay; C14CMC-method; flocculation; radio-labelled carboxymethylcellulase.

XX Humicola insolens; DSM 1800.

XX Key Location/Qualifiers

FT Peptide 1..21  
FT /label= signal

FT Protein 22..305  
/label= endoglucanase

PN EP495258-A1.

XX 22-JUL-1992.

XX 06-NOV-1991; 91EP-00202880.

XX 16-JAN-1991; 91EP-00670006.

XX (PROC) PROCTER & GAMBLE CO.

XX Convents AC, Busch A, Baeck AC;

XX WPI; 1992-243163/30.

XX N-PSDB; AAQ26405.

XX Compact, granular detergent compans. - contain high activity cellulase and softening clay to provide synergistic effect in softening performance.

XX Claim 5; Page 29-30; 38pp; English.

XX The endoglucanase isolated from H. insolens DSM 1800 has Mr 43,000 and high cellulase activity (i.e. removes a minimum of 10% of radio-labelled carboxymethylcellulose (C14CMC)). The enzyme can be used in detergent compositions along with a surface active agent, a builder system and a softening clay. See also AAQ26407. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;

Best Local Similarity 67.1%; Pred. No. 1,4e-80; Indels 10; Gaps 4;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSPVYACDANFOR 60

Db 119 PVAGKTVYVGTSTGSLGSGNHEDINIGGGVGIPIQCTPGFGLPGQRVGGISSNRECCD 178

Db 1 MRSSPLPSAVNALPVLALAA--DGRSTRYWDCKPSGMAKAPVNGVPSGNANFOR 58  
 QY 61 LSDFNVOGCGN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 119  
 Db 59 ITDFPAKSGCBPGVAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 118  
 QY 12C PVAGKTMVQSTSTGDLGSDNFDLAMPGGVGIENGSSQGGIPLGAGVGGISSRQCD 179  
 Db 119 PVAGKTMVQSTSTGDLGSDNFDLAMPGGVGIENGSSQGGIPLGAGVGGISSRQCD 178  
 QY 180 SFPAFLKPGCGMRPFMDPNDPTFTFOQVCPAEIVARSCKRNDSSFP-VFTPPSG 238  
 Db 179 RFPDALKPGCTMRPFMDPNDPTFTFOQVCPAEIVARSCKRNDSSFP-VFTPPSG 238  
 QY 239 NGGTGPTSTAPSGQTS-----PGSGSGCTSQKMAQCGIGPSGCTTCVSGTCTCKLN 292  
 Db 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTCKLN 298  
 QY 293 DYISQCL 299  
 Db 299 DWYHQL 305

## RESULT 9

AA25464  
 ID AA25464 standard; protein; 305 AA.

XX AC AA25464;  
 XX DT 25-MAR-2003 (revised)  
 XX DT 07-JAN-1993 (first entry)  
 XX DE Endoglucanase #1.  
 XX KM CMC-endocase; 43 kD cellulase; monoclonal antibody.  
 XX OS Humicola insolens.  
 XX FH Key Location/Qualifiers  
 XX FT Peptide 1..21  
 XX FT Protein /label=Signal\_peptide  
 XX FT Protein 22..305  
 XX FT Protein /label=Mature\_protein  
 XX PN EP495257-A1.  
 XX PD 22-JUL-1992.  
 XX PF 06-NOV-1991; 91BP-00202879.  
 XX PR 16-JAN-1991; 91EP-00870006.  
 XX PA (PROC) PROCTER & GAMBLE CO.  
 XX PI Baeck AC, Busch A, Ceulemans RAA;  
 XX DR WPI, 1992-243162/30.  
 XX DR N-PSDB; AAQ26380.  
 XX PT Compact, granular detergent compns. - contain high activity cellulase  
 XX PT and softening clay to provide synergistic effect in softening  
 XX PT performance.  
 XX PS Disclosure; Page 20-21; 29pp; English.  
 XX CC The sequences given in AA25464 and AA25466 are endoglucanases which are  
 XX CC immunoreactive with a monoclonal antibody raised against a partially  
 XX CC purified 43 kD cellulase derived from Humicola insolens. These  
 XX CC endoglucanases exhibit a CMC-endocase activity of at least about 50, pref.  
 XX CC at least about 60, imipaticular at least about 90 CMC- endocase units per  
 XX CC mg of total protein. These endoglucanases have molecular weight of  
 XX CC approx. 43 kD. (Updated on 25-MAR-2003 to correct PW field.) (Updated on  
 XX CC 25-MAR-2003 to correct DR field.)

XX Sequence 305 AA:

Query Match 69.9%; Score 1159; DB 2; Length 305;  
 Best Local Similarity 67.1%; Pred. No. 1,4e-80;  
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPVLRITTLAAALPVLASAGSGSTRYWDCKPSGMAKAPVNGVPSGNANFOR 60  
 Db 1 MRSSPLPSAVNALPVLALAA--DGRSTRYWDCKPSGMAKAPVNGVPSGNANFOR 58  
 QY 61 LSDFNVOGCGN-GGSAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 119  
 Db 59 ITDFPAKSGCBPGVAYSCADQTPMAVNDNLAYGPAATSIAGSSBSWCCACALYFTTSG 118  
 QY 120 PVAGKTMVQSTSTGDLGSDNFDLAMPGGVGIENGSSQGGIPLGAGVGGISSRQCD 179  
 Db 119 PVAGKTMVQSTSTGDLGSDNFDLAMPGGVGIENGSSQGGIPLGAGVGGISSRQCD 178  
 QY 180 SFPAFLKPGCGMRPFMDPNDPTFTFOQVCPAEIVARSCKRNDSSFP-VFTPPSG 238  
 Db 179 RFPDALKPGCTMRPFMDPNDPTFTFOQVCPAEIVARSCKRNDSSFP-VFTPPSG 238  
 QY 239 NGGTGPTSTAPSGQTS-----PGSGSGCTSQKMAQCGIGPSGCTTCVSGTCTCKLN 292  
 Db 239 SSPVNGPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTCKLN 298  
 QY 293 DYISQCL 299  
 Db 299 DWYHQL 305

## RESULT 10

AA27968  
 ID AA27968 standard; protein; 305 AA.

XX AC AA27968;  
 XX DT 24-OCT-2003 (revised)  
 XX DT 25-MAR-2003 (revised)  
 XX DT 09-MAR-1993 (first entry)  
 XX DE Endoglucanase enzyme.  
 XX KM Alkaline cellulase; laundry detergent compositions;  
 XX KM fabric colour maintenance.  
 XX OS Humicola insolens; DSM 1800.  
 XX FH Key Location/Qualifiers  
 XX FT Peptide 1..21  
 XX FT Protein /note="signal peptide"  
 XX FT Peptide 22..305  
 XX FT Protein /note="mature peptide"  
 XX PN 3P508358-A1.  
 XX PD 14-OCT-1992.  
 XX PF 07-APR-1992; 92BP-00105956.  
 XX PR 12-APR-1991; 91EP-00870062.  
 XX PA (PROC) PROCTER & GAMBLE CO.  
 XX PI Busch A, Maccorquodale F;  
 XX DR WPI; 1992-341667/42.  
 XX DR N-PSDB; AAQ23934.  
 XX PT Laundry detergent compns. - contg. alkali cellulase and PVP for fabric  
 XX PT colour maintenance upon laundering.

PS Disclosure; Page 16; 23pp; English.

XX  
CC The sequence is that of an endoglucanase (alkaline cellulase) from  
CC Humicola insolens which is used with a polyvinylpyrrolidone (PVP) in a  
CC laundry detergent compsn. The action of the PVP and the endoglucanase in  
CC fabric colour maintenance upon laundering is superior to the sum of the  
CC individual actions of both these ingredients. See also AAR27699. (Updated  
CC on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to  
CC standardise OS field)

XX  
SQ Sequence 305 AA;

Query Match 59.9%; Score 1159; DB 2; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY : KRSTPVLRFTTLAALPLVASAASGSGSTRYDCCPCSCAMPKAAVSGPVYACDANFOR 60  
DB : KRSSPLPSAVVALPVALALAA--DGRSTRYDCCPCSCMAKKAAPVNPVPSCNANFOR 58  
QY 61 LSDFNVOGCV-GGSAVSGADQTPWAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 119  
DB 59 ITDFPAKSGCEPGVAVSCADQTPWAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 118  
QY 120 PVAGKTMVVOGSTSTGSDLSNPFDIAMFGGSGVGFNGSSQFGILPQAYGGISSRDQD 179  
DB 119 PVAGKTMVVOGSTSTGSDLSNPFDIAMFGGSGVGFNGSSQFGILPQAYGGISSRDQD 178  
QY 180 SFPAALPKGCGMRPFQANADNPTTFQVOCPEARIVARSCKRNDSSFP-VTFPPSGG 238  
DB 179 RFPDALPKGCGYRDMFKKADNPSPFRQVQCPALVARTGCRNRDNDGNFPAVOIPISSST 238  
QY 239 NGGTGTPTSTAPSGQTS-----PGSGSGCTSGKMAQCGGIGPSGCTTCVSGTTCQKLN 292  
DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298  
QY 233 EYYSQCL 299  
DB 299 DMYHQCL 305

RESULT 11  
AAR42063  
ID AAR42063 standard; protein; 305 AA.

XX  
AC AAR42063;  
XX  
DT 28-APR-1994 (first entry)  
XX  
DE Endoglucanase enzyme.  
XX  
KW Detergent; fabric; surfactant; softening clay; cellulase;  
KW Humicola insolens; Fusarium oxysporum; endoglucanase.  
XX  
OS Humicola insolens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= sig\_peptide  
FT Protein 22..305  
FT /label= mat\_protein  
XX  
PN AU9211048-A.  
XX  
PD 02-SEP-1993.  
XX  
PF 18-FEB-1992; 52AU-00011048.  
XX  
PR 18-FEB-1992; 52AU-00011048.  
XX  
PA (PROC) PROCTER & GAMBLE CO.  
XX  
PI Corvents AC, Busch A, Baech AC;

XX  
DR WPI; 1993-328419/42.  
DR N-PSDB; AA049941.  
XX  
PT Detergent compsn., esp. for fabrics - contain surfactant, softening clay  
PT and high activity cellulase, partic. from Humicola insolens.  
XX  
PS Claim 5; Page 54-55; 71pp; English.

XX  
CC A new detergent comprises a high activity cellulase in combination with a  
CC softening clay. The cellulase may be an endoglucanase enzyme derived from  
CC Humicola insolens (AA049941) or Fusarium oxysporum (AA049942). The  
CC combination provides a higher than additive softening performance and  
CC excellent colour rejuvenation and whiteness maintenance for fabrics

XX  
SQ Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 KRSTPVLRFTTLAALPLVASAASGSGSTRYDCCPCSCAMPKAAVSGPVYACDANFOR 60  
DB 1 KRSSPLPSAVVALPVALALAA--DGRSTRYDCCPCSCMAKKAAPVNPVPSCNANFOR 58  
QY 61 LSDFNVOGCV-GGSAVSGADQTPWAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 119  
DB 59 ITDFPAKSGCEPGVAVSCADQTPWAVNDNLAYGFATSIAGSSSSWCCACVALTFTSG 118  
QY 120 PVAGKTMVVOGSTSTGSDLSNPFDIAMFGGSGVGFNGSSQFGILPQAYGGISSRDQD 179  
DB 119 PVAGKTMVVOGSTSTGSDLSNPFDIAMFGGSGVGFNGSSQFGILPQAYGGISSRDQD 178  
QY 180 SFPAALPKGCGMRPFQANADNPTTFQVOCPEARIVARSCKRNDSSFP-VTFPPSGG 238  
DB 179 RFPDALPKGCGYRDMFKKADNPSPFRQVQCPALVARTGCRNRDNDGNFPAVOIPISSST 238  
QY 239 NGGTGTPTSTAPSGQTS-----PGSGSGCTSGKMAQCGGIGPSGCTTCVSGTTCQKLN 292  
DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKLN 298  
QY 233 EYYSQCL 299  
DB 299 DMYHQCL 305

RESULT 12  
AAR37150  
ID AAR37150 standard; protein; 305 AA.

XX  
AC AAR37150;  
XX  
DT 25-MAR-2003 (revised)  
DT 25-AUG-1993 (first entry)  
XX  
DE Dye transfer inhibiting compsn. cellulase.  
XX  
KW Detergent; homogeneous endoglucanase component; 43kD cellulase.  
XX  
OS Humicola insolens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "signal peptide"  
FT Peptide 22..305  
FT /note= "mature peptide"  
XX  
PN EP540784-A1.  
XX  
PD 12-MAY-1993.  
XX  
PF 06-NOV-1991; 91EP-00202882.  
XX

PR 36-NOV-1991; 93EP-00202882.  
 XX (PROC ) PROCTER & GAMBLE CO.  
 XX  
 XX Mccorquodale F, Busch A;  
 XX  
 XX WPI: 1993-153688/19.  
 DR N-PSDB: AA041732.  
 XX  
 XX Compn. for inhibiting dye transfer during fabric washing - contains  
 PT peroxidase, hydrogen peroxide, substrate and cellulase, esp. endo-  
 PT glucanase from Humicola insolens.  
 XX  
 XX Disclosure; Page 18-19; 28pp; English.  
 XX  
 XX The sequence is that of a cellulase which is characterised in that the  
 CC cellulase provides at least 10% removal of immobilised radioactive  
 CC labelled carboxymethyl cellulose according to the C14CMC method at 25 x  
 CC 10(-6) by weight of the cellulase protein in the test solution. It can  
 CC be used as part of a compn. for inhibiting dye transfer. (Updated on 25-  
 CC MAR-2003 to correct FN field.)  
 XX  
 XX Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;  
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLTAALPLVVASAASGSGSTRYWDCKRSCAMPKGAASQPVYACDANFOR 60  
 DB 1 MSSPLPSAVVAALPVYALAA--DGRSTRYWDCKRSCGMAKAPVNVQVFSNANFOR 58  
 QY 61 LSDFNVQSGCN-GGSAVSCADOTPMANVDNLAYGPAATSIAGSSSEWCCATLFTTSG 119  
 DB 59 ITDFNKGSGCGPBGVAVSCADOTPMANVDNDFALGPAATSIAGSNAGMCCACVCLTFTSG 118  
 QY 120 PVAGKMWVOSTSTGDLGSGNOFDIAMPGGVGIENGCSGPGGLPGAQYGGISRPDCCD 179  
 DB 119 PVAGKMWVOSTSTGDLGSGNHFDLNIPEGGVGIFDCTQPGGLPGAQYGGISRNEDD 178  
 QY 180 SEPAPLKPQCMRPFQNDNPTFTFQOVQCPAEIVASGCKRNDSSFP-VFTPPSGG 238  
 DB 179 RFPDLKPGCYRPFQNDNPTFTFQOVQCPAEIVARCGCRNDGDFPAVQIPSSST 238  
 QY 239 NGSTGTPTSTAPSGGTS-----PGSGSGTSQKMAQCGGIFSGCTTCVSGTCCOKLN 292  
 DB 239 SSPVQPTSTSTSTSTSTSPVQPTTPSGCTAERMAQCGSGNGSGCTTCVAGSTCTKIN 298  
 QY 293 DYYSQCL 299  
 DB 299 DMYHQCL 305

RESULT 13  
 AAY67388  
 ID AAY67388 standard; protein; 305 AA.  
 XX  
 XX AAY67388;  
 XX  
 XX 25-APR-2000 (first entry)  
 XX  
 XX H. insolens endoglucanase enzyme protein sequence.  
 XX  
 XX Endoglucanase; cellulase enzyme; detergent composition; anion surfactant;  
 KM degradation protection; liquid detergent; long term stability.  
 XX  
 XX Humicola insolens.  
 XX  
 XX EP633311-A1.  
 XX  
 XX 11-JAN-1995.  
 XX  
 XX 28-JUN-1993; 93EP-00870122.

XX  
 XX 28-JUN-1993; 93EP-00870122.  
 XX  
 XX (PROC ) PROCTER & GAMBLE CO.  
 XX  
 XX Herbots IMAU, Jansen MP;  
 XX  
 XX WPI: 1995-038508/06.  
 DR N-PSDB: AAZ60178.  
 XX  
 XX liq. detergent compns. - contg. hydrophobic amine(s) for cellulase  
 PT stabilisation.  
 XX  
 XX Claim 11; Page 18-19; 28pp; English.  
 XX  
 XX This sequence represents the Humicola insolens endoglucanase enzyme  
 CC protein sequence. The enzyme is a cellulase enzyme used in the liquid  
 CC detergent composition of the invention. The detergent comprises an anion  
 CC surfactant, a cellulase enzyme and a cellulase stabilising amount of an  
 CC amine of the formula R1R2R3N; where R1 and R2 are independently H or a C2  
 CC -C9 alkyl chain, and R3 is a C2-C9 alkyl chain or cyclohexyl or  
 CC cyclopentyl or cycloheptyl. The amine prevents the cellulase enzyme from  
 CC becoming degraded. The composition is used as a liquid detergent, and has  
 CC long term stability. It is applicable to the protection of an cellulase  
 CC and can be used in the presence of proteases  
 XX  
 XX Sequence 305 AA;

Query Match 69.9%; Score 1159; DB 2; Length 305;  
 Best Local Similarity 67.1%; Pred. No. 1.4e-80;  
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLTAALPLVVASAASGSGSTRYWDCKRSCAMPKGAASQPVYACDANFOR 60  
 DB 1 MSSPLPSAVVAALPVYALAA--DGRSTRYWDCKRSCGMAKAPVNVQVFSNANFOR 58  
 QY 61 LSDFNVQSGCN-GGSAVSCADOTPMANVDNLAYGPAATSIAGSSSEWCCATLFTTSG 119  
 DB 59 ITDFNKGSGCGPBGVAVSCADOTPMANVDNDFALGPAATSIAGSNAGMCCACVCLTFTSG 118  
 QY 120 PVAGKMWVOSTSTGDLGSGNOFDIAMPGGVGIENGCSGPGGLPGAQYGGISRPDCCD 179  
 DB 119 PVAGKMWVOSTSTGDLGSGNHFDLNIPEGGVGIFDCTQPGGLPGAQYGGISRNEDD 178  
 QY 180 SEPAPLKPQCMRPFQNDNPTFTFQOVQCPAEIVASGCKRNDSSFP-VFTPPSGG 238  
 DB 179 RFPDLKPGCYRPFQNDNPTFTFQOVQCPAEIVARCGCRNDGDFPAVQIPSSST 238  
 QY 239 NGSTGTPTSTAPSGGTS-----PGSGSGTSQKMAQCGGIFSGCTTCVSGTCCOKLN 292  
 DB 239 SSPVQPTSTSTSTSTSPVQPTTPSGCTAERMAQCGSGNGSGCTTCVAGSTCTKIN 298  
 QY 293 DYYSQCL 299  
 DB 299 DMYHQCL 305

RESULT 14  
 AAM01502  
 ID AAM01502 standard; protein; 305 AA.  
 XX  
 XX AAM01502;  
 XX  
 XX 16-OCT-2003 (revised)  
 DT 27-AUG-2003 (revised)  
 DT 26-FEB-1997 (first entry)  
 XX  
 XX 43 kD endoglucanase, EG V.  
 XX  
 XX Detergent composition; cellulase; retaining-type activity;  
 KM catalytic activity; cellulosiose; particulate soil removal;  
 KM colour clarification; cleaning; cellulose-containing fabric;  
 KM cellulohydrolase; endoglucanase.

XX Humicola insolens; DSM 1800.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 XX Peptide 1..21  
 XX /label= signal\_peptide  
 XX Protein 22..305  
 XX /label= mature\_protein  
 XX  
 XX MO9502675-A1.  
 XX  
 XX 26-JAN-1995.  
 XX  
 XX 07-JUL-1994; 94WO-DK00280.  
 XX  
 XX 12-JUL-1993; 93EP-00870131.  
 XX PR 11-OCT-1993; 93DK-00001135.  
 XX  
 XX (NOVO ) NOVO-NORDISK AS.  
 XX (PROC ) PROCTER & GAMBLE CO.  
 XX  
 XX Schuelein M, Convents A, Jeffreys B, Tikhomirov DF,  
 XX WPI; 1995-067325/09.  
 XX  
 XX Detergent compsn. contg. two cellulase components - the first removing  
 XX soil particles and the second capable of colour clarification, useful in  
 XX laundry compans.  
 XX  
 XX Claim 25; Page 66-67; 83pp; English.  
 XX  
 XX Detergent compositions comprising: 1) a first cellulase component having  
 XX retaining-type activity, pref. having a catalytic activity on cellulose  
 XX at pH 8.5 corresponding to kcat of at least 0.01 s<sup>-1</sup>, and capable of  
 XX particulate soil removal; and 2) a second cellulase component having  
 XX multiple domains comprising at least one non-catalytic domain attached  
 XX to a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5  
 XX per 1 mg of cellulase protein higher than 10<-4> IU and being capable of  
 XX colour clarification, where at least one of the cellulase components is a  
 XX single (recombinant) component, are useful for cleaning and colour  
 XX clarification of cellulose-containing fabrics. The second cellulase  
 XX component can be an endoglucanase which is immunoreactive with an  
 XX antibody raised against a highly purified ~43 kD endoglucanase derived  
 XX from Humicola insolens, DSM 1800, and is esp. the present sequence,  
 XX designated EG V. (Updated on 27-AUG-2003 to correct OS field.) (Updated  
 XX on 16-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 305 AA:  
 XX  
 XX Query Match 69.9%; Score 1159; DB 2; Length 305;  
 XX Best Local Similarity 67.1%; Pred. No. 1,4e-80;  
 XX Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;  
 QY 1 MRSPTVLTITLAAALPLVSAASGSGSTRYWDCKKSCAMPAGKAAVSQPYACDANFOR 60  
 DB 1 MRSPTVLTITLAAALPLVSAASGSGSTRYWDCKKSCAMPAGKAAVSQPYACDANFOR 58  
 QY 61 LSDPNVOSGCN-GGSAYSCADOTPMVAVNDLAVFPAATSIAGSSSSWCCACATLFTSG 119  
 DB 59 IDTFPAKSGCEPGVAVVASCADOTPMVAVNDLAVFPAATSIAGSSSSWCCACATLFTSG 118  
 QY 120 PVAGKTMVOSTSTGDLGSSNODIAMPFGGSGVIGENGSSQSGSLPGAQVGGISSSDQCD 179  
 DB 119 PVAGKTMVOSTSTGDLGSSNODIAMPFGGSGVIGENGSSQSGSLPGAQVGGISSSDQCD 178  
 QY 183 SEPAPLKPGCCQMRDWMQUNNFTFPQOVCCPEIYASGGCKRNDSSFP-VFTPPSGG 238  
 DB 179 RFPDAKPGCECTMRDWMQUNNFTFPQOVCCPEIYASGGCKRNDSSFP-VFTPPSGG 238  
 QY 239 NGGTGFTSTAPSGSGTS-----PAGGSGCTSCAKAACCGGIGFSGGCTTCVSGTTQCKIN 292  
 DB 239 SSPVNGFTSTSTSTSTSSPPVQPTTTSCTAFRMAQCGGAGMGSGCTTCVAGSTCTKIN 298

QY 293 DYSOCL 299  
 DB 299 DWTQCL 305  
 RESULT 15  
 AAR88471  
 ID AAR88471 standard; protein; 305 AA.  
 XX  
 XX AAR88471;  
 XX  
 XX 16-OCT-2003 (revised)  
 XX 03-APR-1996 (first entry)  
 XX  
 XX Alkaline endoglucanase Carezyme.  
 XX  
 XX Alkaline endoglucanase; carezyme; cellulase; host cell;  
 XX Fusarium graminearum; heterologous gene expression.  
 XX  
 XX Humicola insolens; DSM 1800.  
 XX  
 XX Key Location/Qualifiers  
 XX Peptide 1..21  
 XX /label= Sig\_peptide  
 XX  
 XX MO9600787-A1.  
 XX  
 XX 15-JUN-1995; 95WO-US007743.  
 XX  
 XX 30-JUN-1994; 94US-00269449.  
 XX PR 15-MAR-1995; 95US-00404678.  
 XX  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX  
 XX Royer JC, Moyer DL, Yoder W, Shuster JR;  
 XX WPI; 1996-077498/08.  
 XX DR N-PSDB; AAT10182.  
 XX  
 XX Non-toxic, non-toxicogenic, non-pathogenic recombinant Fusarium host cell -  
 XX used to produce heterologous proteins, pref. enzymes, hormones, growth  
 XX factors or receptors.  
 XX  
 XX Claim 13; Page 22-24; 38pp; English.  
 XX  
 XX The Humicola insolens DSM 1800 alkaline endoglucanase Carezyme  
 XX (AAR88471), a derivative of a 43 kDa cellulase, is expressed in non-  
 XX toxic, non-toxicogenic, non-pathogenic, protease-deficient recombinant  
 XX Fusarium graminearum ATCC 20334 carrying plasmid pDM151. This plasmid  
 XX includes the Carezyme gene (AAT10182) operably linked to promoter and  
 XX terminator (AAT10184-85) sequences of the Fusarium oxysporum trypsin-like  
 XX protease SP387 gene. Recombinant enzyme is produced at levels of 6.0 g/l  
 XX host cell; over 90% of secreted protein is Carezyme. (Updated on 16-OCT-  
 XX 2003 to standardise OS field)  
 XX  
 XX Sequence 305 AA:  
 XX  
 XX Query Match 69.9%; Score 1159; DB 2; Length 305;  
 XX Best Local Similarity 67.1%; Pred. No. 1,4e-80;  
 XX Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;  
 QY 1 MRSPTVLTITLAAALPLVSAASGSGSTRYWDCKKSCAMPAGKAAVSQPYACDANFOR 60  
 DB 1 MRSPTVLTITLAAALPLVSAASGSGSTRYWDCKKSCAMPAGKAAVSQPYACDANFOR 58  
 QY 61 LSDPNVOSGCN-GGSAYSCADOTPMVAVNDLAVFPAATSIAGSSSSWCCACATLFTSG 119  
 DB 59 IDTFPAKSGCEPGVAVVASCADOTPMVAVNDLAVFPAATSIAGSSSSWCCACATLFTSG 118  
 QY 120 PVAGKTMVOSTSTGDLGSSNODIAMPFGGSGVIGENGSSQSGSLPGAQVGGISSSDQCD 179  
 DB 119 PVAGKTMVOSTSTGDLGSSNODIAMPFGGSGVIGENGSSQSGSLPGAQVGGISSSDQCD 178

```

Db      119 PVAGKMWVQSTSTGDLGSNHPDLNIPGGVGIFDGCTPQPGGLPGQRKYGSSSRNECD 178
Qy      126 SEPAFLPBGQWRPFQONADNPTFTFOYOCBAIIVARSCQRNDSSFP-VFTPPSGG 238
      179 RFPDALPBGQWRPFQONADNPSFSFQYCCPAELVARTGCRNDGNGFPAVOIPSSST 238
Qy      239 NGGCTPTSTAPSGGOTS-----PQGGSGCTSKMAQCGGIGFSGCTTCVSGTTCKM 232
Db      239 SSPVNOPTISTISTSTSSPPVQPTTPSGCTAERMAQCGANGSGCTTCVAGSTCTKIN 298
Qy      293 DYVSQCL 299
      299 LMYHQL 305

```

Search completed: July 7, 2004, 08:49:38  
 Job time : 61 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 22 Seconds

(without alignments)  
701.644 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSFPUKRTIAALPLVAS.....TTCVSGCTCKKNDYISQCL 299

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Parents AA:\*  
2: /cgm2\_6/prodata/2/1aa/5A.COMB.pep:\*  
3: /cgm2\_6/prodata/2/1aa/5B.COMB.pep:\*  
4: /cgm2\_6/prodata/2/1aa/6A.COMB.pep:\*  
5: /cgm2\_6/prodata/2/1aa/6B.COMB.pep:\*  
6: /cgm2\_6/prodata/2/1aa/PCUTS.COMB.pep:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1659	100.0	299	2	US-08-872-437-2
2	1659	100.0	299	3	US-08-651-136C-12
3	1659	100.0	299	4	US-09-229-911A-6
4	1178	71.0	305	3	US-09-230-222-1
5	1159	69.9	305	1	US-08-090-013-2
6	1159	69.9	305	1	US-08-081-328-2
7	1159	69.9	305	1	US-08-232-249-2
8	1159	69.9	305	2	US-08-921-426-8
9	1159	69.9	305	2	US-08-833-642A-2
10	1159	69.9	305	2	US-08-140-008A-4
11	1159	69.9	305	2	US-08-836-340-1
12	1159	69.9	305	2	US-08-389-423-2
13	1159	69.9	305	3	US-08-816-915-8
14	1159	69.9	305	3	US-09-189-060B-56
15	1159	69.9	305	4	US-09-230-665-2
16	1159	69.9	305	4	US-09-189-028-2
17	1159	69.9	305	5	PCT-US35-07743-8
18	1142.5	68.9	306	3	US-09-189-060B-66
19	1127	67.9	286	3	US-09-254-733-3
20	1125	67.8	285	4	US-09-230-225B-6
21	1112	67.0	284	1	US-08-411-777-10
22	1112	67.0	284	3	US-09-057-088-10
23	1107.5	66.8	306	3	US-09-189-060B-68
24	1076.5	64.9	308	3	US-09-189-060B-70
25	1076.5	64.9	304	3	US-09-189-060B-72
26	1045	63.0	307	3	US-09-189-060B-74
27	1039	62.6	235	3	US-09-329-350-31

28	1021	61.5	308	3	US-08-651-136C-6	Sequence 6, Appl1
29	1021	61.5	308	4	US-09-229-911A-6	Sequence 6, Appl1
30	994.5	59.9	297	3	US-08-651-136C-4	Sequence 4, Appl1
31	994.5	59.9	297	4	US-09-229-911A-4	Sequence 4, Appl1
32	968.5	58.4	234	4	US-09-230-665-6	Sequence 6, Appl1
33	962	58.0	310	3	US-08-651-136C-22	Sequence 22, Appl1
34	962	58.0	310	4	US-09-229-911A-22	Sequence 22, Appl1
35	951	57.3	295	3	US-08-651-136C-8	Sequence 8, Appl1
36	951	57.3	295	4	US-09-229-911A-8	Sequence 8, Appl1
37	939.5	56.6	376	1	US-08-090-013-4	Sequence 4, Appl1
38	939.5	56.6	376	1	US-08-081-328-4	Sequence 4, Appl1
39	939.5	56.6	376	1	US-08-232-249-4	Sequence 4, Appl1
40	939.5	56.6	376	2	US-08-833-642A-4	Sequence 4, Appl1
41	939.5	56.6	376	2	US-08-389-423-4	Sequence 4, Appl1
42	939.5	56.6	376	4	US-09-189-028-4	Sequence 4, Appl1
43	934.5	56.3	376	4	US-09-230-665-4	Sequence 4, Appl1
44	923.5	55.7	294	3	US-08-651-136C-24	Sequence 24, Appl1
45	923.5	55.7	294	4	US-09-229-911A-24	Sequence 24, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-872-437-2  
Sequence 2, Application US/08872437

Patent No. 5958082

GENERAL INFORMATION:

APPLICANT: Lund, Henrik

APPLICANT: Kallum, Lisbeth

TITLE OF INVENTION: Garments with Considerable Variation In

FILE REFERENCE: 4888, 200- US

CURRENT APPLICATION NUMBER: US/08/872, 437

CURRENT FILING DATE: 1997-06-10

EARLIER APPLICATION NUMBER: 1276/96

EARLIER FILING DATE: 1996-11-13

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 299

TYPE: PRT

ORGANISM: Thielavia terrestris

US-08-872-437-2

Query Match 100.0%; Score 1659; DB 2; Length 299;

Best Local Similarity 100.0%; Pred. No. 2, 5e-127; Mismatches 0; Indels 0; Gaps 0;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRSFVLRRTIAALPLVASAAGSGGSTRYMDCKPSCAMPGRAAVSQPVYACDANFOR	60
DB	1	MRSFVLRRTIAALPLVASAAGSGGSTRYMDCKPSCAMPGRAAVSQPVYACDANFOR	60
QY	61	ISDFVSGNGGSAVSCADQTPWAVNDNLAYGFAATSIAGSSSSMCCAYLTFTSGP	120
DB	61	ISDFVSGNGGSAVSCADQTPWAVNDNLAYGFAATSIAGSSSSMCCAYLTFTSGP	120
QY	121	VAGKTMVQSTNGSGDLSNCFDIAMPGGVGFNCCSQRFGI,PGAQYGGISRPQDCS	180
DB	121	VAGKTMVQSTNGSGDLSNCFDIAMPGGVGFNCCSQRFGI,PGAQYGGISRPQDCS	180
QY	181	FPAPLKPCCQMRDMFQNDNPTFTFOOVCPAEIYVARSQKRRDSSFFVTPPGANG	240
DB	181	FPAPLKPCCQMRDMFQNDNPTFTFOOVCPAEIYVARSQKRRDSSFFVTPPGANG	240
QY	241	GTGFTSTAPSGSGTSPGSGSGCTSGVWACCGGIGSGCTTCUSGTTCCQLNDYISQCL	299
DB	241	GTGFTSTAPSGSGTSPGSGSGCTSGVWACCGGIGSGCTTCUSGTTCCQLNDYISQCL	299

RESULT 2  
US-08-651-136C-12  
Sequence 12, Application US/08651136C

Patent No. 6001639  
GENERAL INFORMATION:  
APPLICANT: Schuilein, Martin  
APPLICANT: Andersen, Lene N.  
APPLICANT: Lassen, Soren F.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Lange, Lene  
APPLICANT: Nielsen, Ruby I.  
APPLICANT: Ihara, Michiko  
APPLICANT: Takagi, Shinobu  
TITLE OF INVENTION: No. 6001639e1 Endoglucanases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60016390 No. 6001639e1 disk of No. 6001639th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136C  
FILING DATE: 21-May-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366-200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-651-136C-12

Query Match 100.0%; Score 1659; DB 3; Length 299;

Best Local Similarity 100.0%; Pred. No. 2.5e-127; Mismatches 0; Indels 0; Gaps 0;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
DB 1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
QY 61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
DB 61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
QY 121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
DB 121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
QY 181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
DB 181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
QY 241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299  
241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299  
DB 241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299

## RESULT 3

US-09-229-911A-12  
Sequence 12, Application US/09229911A  
Patent No. 6387690  
GENERAL INFORMATION:

APPLICANT: Schuilein, Martin  
Andersen, Lene N.  
Lassen, Soren F.  
Kauppinen, Markus S.  
Lange, Lene  
Nielsen, Ruby I.  
Ihara, Michiko  
Takagi, Shinobu  
TITLE OF INVENTION: No. 6387690e1 Endoglucanases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 63876900 No. 6387690e1 disk of No. 6387690th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/229,911A  
FILING DATE: 13-Jan-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/651,136  
FILING DATE: 21-May-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366-200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-229-911A-12

Query Match 100.0%; Score 1659; DB 4; Length 299;

Best Local Similarity 100.0%; Pred. No. 2.5e-127; Mismatches 0; Indels 0; Gaps 0;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
DB 1 MRSFVLTTLAALPLVASAASGSGSTRYMDCKPSCAMPKAAVSQPVYACDANFOR 60  
QY 61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
DB 61 LSDFNVOGCGNGSAYSCADQTPMAVNDNLAYGPAATSIAGSSSSWCCACVLTFTSGP 120  
QY 121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
DB 121 VAGKTMVOSTSTGDLASNOFDIAMPGGVGFPGCSCSOPGGLPAQYGGISRRQCD 180  
QY 181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
DB 181 FPAALPGCGQWRPFQNDNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPGSGNG 240  
QY 241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299  
241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299  
DB 241 GTGTPSTAPSGGTSFGGSGCTSQMAQCGGIGSGCTTGVSTGCKLNDYISQCL 299

## RESULT 4

US-09-230-222-1  
Sequence 1, Application US/09230222A

Query Match	71.0%;	Score 1178;	DB 3;	Length 305;
Best Local Similarity	68.1%;	Pred. No. 3.1e-88;		
Matches 209;	Conservative 44;	Mismatches 44;	Indels 10;	Gaps 4;

RESULT 5  
CS-08-090-013-2

1 GENERAL INFORMATION:  
2 APPLICANT: CONVENTS, ANDRE C  
3 APPLICANT: BUSCH, ALFRED  
4 APPLICANT: BAECK, ANDRE C  
5 TITLE OF INVENTION: DETERGENT COMPOSITION WITH HIGH ACTIVITY  
6 TITLE OF INVENTION: CELLULASE AND SOFTENING CLAYS  
7 NUMBER OF SEQUENCES: 4  
8 CORRESPONDENCE ADDRESS:  
9 ADDRESSEE: THE PROCTER & GAMBLE COMPANY  
10 STREET: 5299 SPRING GROVE AVENUE  
11 CITY: CINCINNATI  
12 STATE: OHIO  
13 COUNTRY: USA  
14 ZIP: 45217  
15 COMPUTER READABLE FORM:  
16 MEDIUM TYPE: Floppy disk  
17 COMPUTER: IBM PC compatible

Query Match	69.8%;	Score 1159;	DB 1;	Length 305;
Best Local Similarity	67.1%;	Pred. No. 1,1e-86;		
Matches 206;	Conservative 44;	Mismatches 47;	Indels 10;	Gaps 4;

```

QY 1 MRSPLVLTATLAAALPMAAASGGGSTRMWCCKRSCAPMGAAASQPVACANROR 60
Db 1 MRSPLVLSAVVAAALPVLALAA--DGRSTRWOCCKRSGMAKAKAPMNOVPSCANROR 58
QY 61 LSPDFVNSGSCN--GGSAYS CADQTPMAVANDMLAGFAATSIAGGSESSMCCAYALFTFSG 119
Db 59 ITLPDAKSCGEGGVAIXSADQTPMAVANDPDLAAGFAATSIAGSNBAGMCCAYELFTFSG 118
QY 120 PVAGKTMVVOSTTSGDSLNSNOPLDAMPGGGCGVIGENGSSQPGCLPQAQYGGISRRDCD 179
Db 119 PVAGKTMVVOSTTSGDSLNSHFDLNTINGGGGCGVIGFGCTPGFGCLPQARGGISRRNCD 178
QY 160 SFAPLAPGCCCKRRPDMQNAANDPTEFQVOCALETVARSCKCKNDSSSFP--VTPPSCG 236
Db 179 RFPDAKPCQCYRRDPMKNDNPSFSFPQVOCALETVARTGRNDGDNFPAQVPPSSST 238
QY 239 NGGNGPRTSTARGSGORS-----PQSGSGCTSOXMAOCGGGIGSGCTTGVSGCTCOJLAN 292
Db 239 SSPPNPQRTSTSTSTSTSSPPVQPTTPSGGCTARMAQCGGNGMSGCTTGVASGCTCYKIN 298
QY 293 DYSYSQL 299
Db 299 DMYHQL 305

```

## RESULT 6

Sequence 2, Application US/08081328  
Patent No. 5520838  
GENERAL INFORMATION:  
APPLICANT: BAECK, ANDRE C.  
APPLICANT: CEULEMANS, RAPHAEL, ANGELINE A.  
APPLICANT: BUSCH, ALFRED (NMN)  
TITLE OF INVENTION: COMPACT DETERGENT COMPOSITIONS WITH  
TITLE OF INVENTION: HIGH ACTIVITY CELLULASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE PROCTER & GAMBLE COMPANY  
STREET: 11810 East Miami River Road  
CITY: CINCINNATI  
STATE: OHIO  
COUNTRY: USA  
ZIP: 45253-8707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091.328  
FILING DATE: 11/19/93  
CLASSIFICATION: 252  
ATTORNEY/AGENT INFORMATION:  
NAME: ZERRY, KIM WILLIAM  
REGISTRATION NUMBER: 32,323  
REFERENCE/DOCKET NUMBER: CM56M  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-627-2885  
TELEFAX: 513-627-0318  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-081-328-2

Query Match 69.9%; Score 1159; DB 1; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1.1e-86;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 60  
DB 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 58  
QY 61 LSEFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGAATSIAGSSBSWCCACALFTTSG 119  
DB 59 ITDFPAKSGCEPQVAVASCDQTPMAVNDNLAYGAATSIAGSSBSWCCACALFTTSG 118  
QY 120 PVAKTMYVOSTSTGDLGSGNOFDIAMPGGVGIENGCSQFGSLPGAQYGGISRDQCD 179  
DB 119 PVAKKMYVOSTSTGDLGSGNHEDLNPGGVGIENGCSQFGSLPGAQYGGISRDQCD 178  
QY 180 SFPAALXGCOMFDFPQADNPTFTPOVCPAEIVASGCKRNDSSP-VFTPSGG 238  
DB 179 RFPDALXGCOMFDFPQADNPTFTPOVCPAEIVASGCKRNDSSP-VFTPSGG 238  
QY 239 NGGTGPTSTAPSGGTS-----PGSGSGCTSQMAACGGIGPSGCTTCVSGTTCOKLN 292  
DB 239 SSPVNOPTSTSTSTSTSPVQPTPSGCTAERMAACGGNGSGCTTCVSGTTCOKLN 298  
QY 293 DYSGCL 299  
DB 293 DWTGCL 305

RESULT 7  
US-08-232-249-2

Sequence 2, Application US/08232249  
Patent No. 5610129  
GENERAL INFORMATION:  
APPLICANT: MCCORDUDALE, FIVLAY (NMN)  
APPLICANT: BUSCH, ALFRED (NMN)  
TITLE OF INVENTION: DYE TRANSFER INHIBITING COMPOSITIONS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: THE PROCTER & GAMBLE COMPANY  
STREET: 5299 SPRING GROVE AVENUE  
CITY: CINCINNATI  
STATE: OHIO  
COUNTRY: USA  
ZIP: 45217  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232.249  
FILING DATE: 05-MAY-1994  
CLASSIFICATION: 510  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91202862.6  
FILING DATE: 06-NOV-91  
ATTORNEY/AGENT INFORMATION:  
NAME: ALLEN, GEORGE W.  
REGISTRATION NUMBER: 26,143  
REFERENCE/DOCKET NUMBER: CH-395  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-627-5946  
TELEFAX: 513-627-8118  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-232-249-2

Query Match 69.9%; Score 1159; DB 1; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1.1e-86;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 60  
DB 1 MRSPTVLTAAALPLVASAASGSGSTRYWDCKPCSCAMPKAAVOPVYACDANFOR 58  
QY 61 LSEFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGAATSIAGSSBSWCCACALFTTSG 119  
DB 59 ITDFPAKSGCEPQVAVASCDQTPMAVNDNLAYGAATSIAGSSBSWCCACALFTTSG 118  
QY 120 PVAKTMYVOSTSTGDLGSGNOFDIAMPGGVGIENGCSQFGSLPGAQYGGISRDQCD 179  
DB 119 PVAKKMYVOSTSTGDLGSGNHEDLNPGGVGIENGCSQFGSLPGAQYGGISRDQCD 178  
QY 180 SFPAALXGCOMFDFPQADNPTFTPOVCPAEIVASGCKRNDSSP-VFTPSGG 238  
DB 179 RFPDALXGCOMFDFPQADNPTFTPOVCPAEIVASGCKRNDSSP-VFTPSGG 238  
QY 239 NGGTGPTSTAPSGGTS-----PGSGSGCTSQMAACGGIGPSGCTTCVSGTTCOKLN 292  
DB 239 SSPVNOPTSTSTSTSTSPVQPTPSGCTAERMAACGGNGSGCTTCVSGTTCOKLN 298  
QY 293 DYSGCL 299  
DB 293 DWTGCL 305

RESULT 8  
US-08-921-426-8

Sequence 8, Application US/08921426  
Patent No. 5837847  
GENERAL INFORMATION:  
APPLICANT: ROYER, JOHN C  
APPLICANT: MOYER, DONNA L  
APPLICANT: YODER, WENDY R  
TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 5837847 No. 5837847/disk of No. 5837847th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/921,426  
FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/456,433  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: US 08/404,678  
FILING DATE: 15-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Agis Dr., Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 4216, 010-JS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-921-426-8

Query Match 69.9%; Score 1159; DB 2; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1,1e-86;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MESTPVLRTIAALALPVASAGSGQSTRYWDCKKPCSCAMPKAAVSPVACDANFOR 60  
DB 1 MRSSPLPSAVVAALPVLAAL--DGRSTRYWDCKKPCSCAMPKAAVSPVACDANFOR 58  
QY 61 LSDPNVSGCN-GGSAYSCADQTPMAVNDLAFATSIAGSSSSWCCACVALTFTSG 119  
DB 59 ITDPDAKSGCEPGVAVSCADQTPMAVNDLAFATSIAGSSSSWCCACVALTFTSG 118  
QY 120 PVAKTMYVVGSTGSDLSNQFDIAMPGGVGIIFNGCSSQGGIPLAQYGGISSRQCD 179  
DB 119 PVAGKMYVVGSTGSDLSNQFDIAMPGGVGIIFNGCSSQGGIPLAQYGGISSRQCD 178  
QY 180 SFPAKPKRCOMRFDMFONADNPTFTFOVCPAEIYVARSCKRNDSPF-VTPPSGG 238  
DB 179 RFPDAKPKCGTMRFDKADNPFSPFQVCPAEIYVARSCKRNDSPFPAVOIPSSST 238  
QY 239 NGGTGPTSTAPSGSQT-----PGGSGSCTSQRMAQCGGIGFSGCTTCVSGTTCOKIN 292  
DB 239 SSPVQPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGIGFSGCTTCVSGTTCOKIN 298  
QY 293 DYVSQCL 299  
DB 299 DWYHQCL 305

RESULT 9  
US-08-833-642A-2  
Sequence 2, Application US/08833642A  
Patent No. 5883066  
GENERAL INFORMATION:  
APPLICANT: Ivan M. A. J. Herbots et al.  
TITLE OF INVENTION: Liquid Detergent Compositions  
CONTAINING CELLULOSE AND AMINE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jackie Ann Zurcher  
ADDRESS: 255 E. Fifth Street  
STREET: 1900 Chemed Center  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/833,642A  
FILING DATE: April 8, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Zurcher, J. A.  
REGISTRATION NUMBER: P42, 251  
REFERENCE/DOCKET NUMBER: CM551C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 977-8377  
TELEFAX: (513) 977-8141  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-833-642A-2

Query Match 69.9%; Score 1159; DB 2; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1,1e-86;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MESTPVLRTIAALALPVASAGSGQSTRYWDCKKPCSCAMPKAAVSPVACDANFOR 60  
DB 1 MRSSPLPSAVVAALPVLAAL--DGRSTRYWDCKKPCSCAMPKAAVSPVACDANFOR 58  
QY 61 LSDPNVSGCN-GGSAYSCADQTPMAVNDLAFATSIAGSSSSWCCACVALTFTSG 119  
DB 59 ITDPDAKSGCEPGVAVSCADQTPMAVNDLAFATSIAGSSSSWCCACVALTFTSG 118  
QY 120 PVAKTMYVVGSTGSDLSNQFDIAMPGGVGIIFNGCSSQGGIPLAQYGGISSRQCD 179  
DB 119 PVAGKMYVVGSTGSDLSNQFDIAMPGGVGIIFNGCSSQGGIPLAQYGGISSRQCD 178  
QY 180 SFPAKPKRCOMRFDMFONADNPTFTFOVCPAEIYVARSCKRNDSPF-VTPPSGG 238  
DB 179 RFPDAKPKCGTMRFDKADNPFSPFQVCPAEIYVARSCKRNDSPFPAVOIPSSST 238  
QY 239 NGGTGPTSTAPSGSQT-----PGGSGSCTSQRMAQCGGIGFSGCTTCVSGTTCOKIN 292  
DB 239 SSPVQPTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGIGFSGCTTCVSGTTCOKIN 298  
QY 293 DYVSQCL 299  
DB 299 DWYHQCL 305

RESULT 10  
US-08-140-008A-4  
Sequence 4, Application US/0814008A  
Patent No. 5914306  
GENERAL INFORMATION:  
APPLICANT: SVENDSEN, Allan  
APPLICANT: VON DER OSTEN, Claus  
APPLICANT: CLAUSEN, Id Groth  
APPLICANT: PATKAR, Shankant Anant  
APPLICANT: BORCH, Kim  
TITLE OF INVENTION: STABILIZED ENZYMES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5914306 No. 5914306disk of No. 5914306th America  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



RESULT 13  
US-08-816-915-8  
Sequence 8, Application US/08816915  
Patent No. 6060305  
GENERAL INFORMATION:  
APPLICANT: Royer, John C  
APPLICANT: Royer, Donna L  
APPLICANT: Yoder, Wendy T  
APPLICANT: Shuster, Jeffrey R  
TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC  
TITLE OF INVENTION: FUSARIUM EXPRESSION SYSTEM  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60603050 No. 6060305disk of No. 6060305th America, Inc  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,915  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Agilis Dr., Cheryl H.

```

RESULT 14
US-09-189-060B-56
Sequence 56, Application US/09189060B
Patent No. 6270968
GENERAL INFORMATION:
APPLICANT: Dalboege, Henrik
APPLICANT: Sandal, Thomas
APPLICANT: Kaupinen, Marcus
APPLICANT: Borge, Diderichsen
TITLE OF INVENTION: Method of Providing No. 6270968e1 DNA Sequences
FILE REFERENCE: 4772.204-US
CURRENT APPLICATION NUMBER: US/09/189,360B
CURRENT FILING DATE: 1998-11-10
PRIORITY APPLICATION NUMBER: PCT/DK97/00216
PRIORITY FILING DATE: 1997-05-12
NUMBER OF SEQ ID NOS: 74
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 56
LENGTH: 305
TYPE: PR1
ORGANISM: Humicola insolens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: (1)...(21)
US-09-189-060B-56

```

Query Match	69.9%;	Score 1159;	DB 3;	Length 305;
Best Local Similarity	67.1%;	Pred. No. 1,1e-86;		
Matches 206;	Conservative 44;	Mismatches 47;	Indels 10;	Gaps
QY	1	MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKCKSCAMPQKAAVSQPYACDANFQR	60	

Search completed: July 7, 2004, 08:46:47  
Job time : 23 secs

```
Db 1 MRSSPLTSAVVAALPVIALAA--DGRSTRWDCCKNSCKMAKAPVNPVFSCKNPFQR 58
Qy 61 ISDPNVQSGCN-GGSAYSACADQTPMAVNDNNAIYGAATSIAGSSSSCCACVLTFTSG 119
Db 59 ITDFDAKSGCBGPVAVYSCADQTPMAVNDNNAIYGAATSIAGSNEAGWCCACVLTFTSG 118
Qy 120 PVAGKTMVYQSTSTGGDLGASNOPDILAMPQGGVGIENGCSQFQGLPGAQYGGISSRDQCD 179
Db 119 PVAGKTMVYQSTSTGGDLGASNOPDILAMPQGGVGIENGCSQFQGLPGAQYGGISSRDQCD 178
Qy 180 SFPAPLKEGCKMREDFWONADNPTFTFOVQCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDALKEGCKMREDFWONADNPTFTFOVQCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Qy 239 NGGTGTPSTAPGSGQIS-----PGGSGCTSQMAQCGGIGFSGCTTCVSGTTCCKIN 292
Db 239 SSPVNOPTSTSTSTSTSSPVQPTTPSGCTABRMWQCGNGMSGCTTCVAGSTCTKIN 298
Qy 293 DYSQCL 299
Db 299 DMYHQL 305
```

```
RESULT 15
US-09-230-665-2
; Sequence 2, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
; APPLICANT: Boyer, Stanton I
; TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
; TITLE OF INVENTION: Components, with and without a Cellulose-Binding Domain
; FILE REFERENCE: 6191 Sequence listing (8 Sequences)
; Patent No. 6322595
; CURRENT APPLICATION NUMBER: US/09/230,665
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/023,125
; EARLIER FILING DATE: 1996-07-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Humicola insolens
US-09-230-665-2
```

Query Match 69.9%; Score 1159; DB 4; Length 305;  
Best Local Similarity 67.1%; Pred. No. 1,1e-86;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```
Qy 1 MRSTPVLRTTLAAALPIVAAASGSGSTRYWDCCCKPSCAMPGKAAVSOPVYACDANFQR 60
Db 1 MRSSPLTSAVVAALPVIALAA--DGRSTRWDCCKNSCKMAKAPVNPVFSCKNPFQR 58
Qy 61 LSDENVQSGCN-GGSAYSACADQTPMAVNDNNAIYGAATSIAGSSSSWCCACVLTFTSG 119
Db 59 ITDFDAKSGCBGPVAVYSCADQTPMAVNDNNAIYGAATSIAGSNEAGWCCACVLTFTSG 118
Qy 120 PVAGKTMVYQSTSTGGDLGASNOPDILAMPQGGVGIENGCSQFQGLPGAQYGGISSRDQCD 179
Db 119 PVAGKTMVYQSTSTGGDLGASNOPDILAMPQGGVGIENGCSQFQGLPGAQYGGISSRDQCD 178
Qy 180 SFPAPLKEGCKMREDFWONADNPTFTFOVQCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDALKEGCKMREDFWONADNPTFTFOVQCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Qy 239 NGGTGTPSTAPGSGQIS-----PGGSGCTSQMAQCGGIGFSGCTTCVSGTTCCKIN 292
Db 239 SSPVNOPTSTSTSTSTSSPVQPTTPSGCTABRMWQCGNGMSGCTTCVAGSTCTKIN 298
Qy 293 DYSQCL 299
Db 299 DMYHQL 305
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:10 ; Search time 51 Seconds  
(without alignments)  
1824.978 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSFVPTRTTAAALPIVAS.....TTCVSGCTCKKNDYSSQL 299

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_5/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_5/prodata/1/pubpaa/PCF\_NEW\_PUB.pep:\*  
3: /cgn2\_5/prodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_5/prodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_5/prodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_5/prodata/1/pubpaa/PCFUS\_PUBCOMB.pep:\*  
7: /cgn2\_5/prodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_5/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_5/prodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_5/prodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_5/prodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_5/prodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_5/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_5/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_5/prodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_5/prodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_5/prodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_5/prodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1659	100.0	299 14 US-10-007-521-12	Sequence 12, Appl
2	1216	73.3	296 15 US-10-369-493-4204	Sequence 4204, Ap
3	1159	69.9	305 9 US-09-735-787-2	Sequence 2, Appl
4	1159	69.9	305 14 US-10-138-870-2	Sequence 2, Appl
5	1122	67.6	201 10 US-09-261-329-5	Sequence 5, Appl
6	1039	62.6	235 8 US-08-841-636A-31	Sequence 31, Appl
7	1021	61.5	308 14 US-10-007-521-6	Sequence 6, Appl
8	994.5	59.9	297 14 US-10-007-521-4	Sequence 4, Appl
9	962	58.0	310 14 US-10-007-521-22	Sequence 22, Appl
10	951	57.3	295 14 US-10-007-521-8	Sequence 8, Appl
11	939.5	56.6	376 9 US-09-735-787-4	Sequence 4, Appl
12	939.5	56.6	376 14 US-10-138-870-4	Sequence 4, Appl
13	938	55.7	201 10 US-09-261-329-4	Sequence 4, Appl
14	923.5	55.7	294 14 US-10-007-521-24	Sequence 24, Appl
15	910.5	54.9	349 14 US-10-007-521-10	Sequence 10, Appl

16	887.5	53.5	202 10 US-09-261-329-1	Sequence 1, Appl
17	887	53.5	298 14 US-10-007-521-18	Sequence 18, Appl
18	884	53.3	293 14 US-10-007-521-20	Sequence 20, Appl
19	819	49.4	203 10 US-09-261-329-6	Sequence 6, Appl
20	801.5	48.3	223 12 US-10-432-290-38	Sequence 38, Appl
21	796.5	47.8	225 14 US-10-007-521-2	Sequence 2, Appl
22	792.5	47.8	228 12 US-10-432-290-24	Sequence 24, Appl
23	777	46.8	205 10 US-09-261-329-7	Sequence 7, Appl
24	769.5	46.4	338 12 US-10-416-328-4	Sequence 4, Appl
25	769.5	46.4	338 12 US-10-432-290-7	Sequence 7, Appl
26	769.5	46.4	387 12 US-10-416-328-5	Sequence 5, Appl
27	769.5	46.4	387 12 US-10-432-290-9	Sequence 9, Appl
28	766.5	46.2	245 12 US-10-432-290-29	Sequence 29, Appl
29	757.5	45.7	366 12 US-10-416-328-2	Sequence 2, Appl
30	757.5	45.7	366 12 US-10-432-290-3	Sequence 3, Appl
31	756.5	45.6	338 12 US-10-416-328-1	Sequence 1, Appl
32	756.5	45.6	338 12 US-10-432-290-1	Sequence 1, Appl
33	752.5	45.4	202 10 US-09-261-329-3	Sequence 3, Appl
34	745	44.9	360 12 US-10-416-328-3	Sequence 3, Appl
35	745	44.9	360 12 US-10-432-290-5	Sequence 5, Appl
36	725	43.7	222 14 US-10-007-521-14	Sequence 14, Appl
37	722.5	43.6	202 10 US-09-261-329-2	Sequence 2, Appl
38	718	43.3	203 10 US-09-261-329-9	Sequence 9, Appl
39	688.5	41.5	226 14 US-10-007-521-16	Sequence 16, Appl
40	686.5	41.4	346 12 US-10-416-328-6	Sequence 6, Appl
41	686.5	41.4	346 12 US-10-432-290-11	Sequence 11, Appl
42	672	40.5	203 10 US-09-261-329-8	Sequence 8, Appl
43	587	35.4	138 14 US-10-007-521-26	Sequence 26, Appl
44	496	29.9	235 10 US-09-261-329-10	Sequence 10, Appl
45	419	25.3	211 10 US-09-261-329-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-10-007-521-12  
Sequence 12, Application US/10007521  
Publication No. US20030054539A1  
GENERAL INFORMATION:  
APPLICANT: Schultein, Martin  
Andersen, Lene N.  
Lassen, Soren F.  
Kauppinen, Markus S.  
Lange, Lene  
Nielsen, Ruby I.  
Ihara, Michiko  
Takagi, Shinobu  
TITLE OF INVENTION: No. US20030054539A1e1 Endoglucaases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. US20030054539A1o No. US20030054539A1disk of No. US20030054539A1  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/007,521  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136  
FILING DATE: 21-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366,200-US

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-10-007-521-12

## Query Match

Best Local Similarity 100.0%; Score 1659; DB 14; Length 299;  
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTPVLRTTLAALPLVASAASGSGSTRYMDCKPCAMPKGAASVPYACDANFOR 60  
DB 1 MSTPVLRTTLAALPLVASAASGSGSTRYMDCKPCAMPKGAASVPYACDANFOR 60  
QY 61 LSDFWVSGGCGSAAVSCADQTPMAVNDNLAYGFAATSIAGSSSSMCACVALTPTSGP 120  
DB 61 LSDFWVSGGCGSAAVSCADQTPMAVNDNLAYGFAATSIAGSSSSMCACVALTPTSGP 120  
QY 121 VAGKTMVVGSTGTGDLGSGNCPDIAMPGGVGIIFNGCSSQFGSLPGAQYGGIISRDQDS 180  
DB 121 VAGKTMVVGSTGTGDLGSGNCPDIAMPGGVGIIFNGCSSQFGSLPGAQYGGIISRDQDS 180  
QY 181 PAPLKPCCQMRPFQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFPPTSGNG 240  
DB 181 PAPLKPCCQMRPFQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFPPTSGNG 240  
QY 241 GGTGTGAPSGSGTSPGGSGCTSOXMAOQGGIGSGCTTCVSGTTCQKLDNYISQL 299  
DB 241 GGTGTGAPSGSGTSPGGSGCTSOXMAOQGGIGSGCTTCVSGTTCQKLDNYISQL 299

## RESULT 2

US-10-369-493-4204  
Sequence 4204, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10152521B  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 4204  
LENGTH: 299  
TYPE: PRT  
ORGANISM: Neurospora crassa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)-(296)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-4204

## Query Match

Best Local Similarity 73.3%; Score 1216; DB 15; Length 296;  
Matches 215; Conservative 27; Mismatches 36; Indels 24; Gaps 3;

QY 15 LPLVVAASGSGSTRYMDCKPCAMPKGAASVPYACDANFORLSDFWVSGGCGS 74  
DB 1 LPLVVAASGSGSTRYMDCKPCAMPKGAASVPYACDANFORLSDFWVSGGCGS 60

QY 75 AYSCADQTPMAVNDNLAYGFAATSIAGSSSSMCACV-----ALTF 116  
DB 61 AYTCANNSPMAVNDNLAYGFAATSIAGSSSSMCACVAYSLQLLGITRRLTFAALTF 120  
QY 117 TSGPVAAGTMVVGSTGTGDLGSGNCPDIAMPGGVGIIFNGCSSQFGSLPGAQYGGIISRD 176  
DB 121 TSGPVAAGTMVVGSTGTGDLGSGNCPDIAMPGGVGIIFNGCSSQFGSLPGAQYGGIISRD 180  
QY 177 QCDSPAPLKPCCQMRPFQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFPPTSPS 236  
DB 181 QCDSPAPLKPCCQMRPFQNDNPTFTFOVQCPAIVARSGCKRNDSSPFVFPPTSPS 240  
QY 237 GANSGTGTGAPSGSGTSPGGSGCTSOXMAOQGGIGSGCTTCVSGTTCQKLDNYIS 296  
DB 241 GGSNSSTPT-TPSS-----GGSGCTADKXKAOQGGSGNSGCTTCSGCTCKTINDYH 294  
QY 297 QC 298  
DB 295 QC 296

## RESULT 3

US-09-735-787-2  
Sequence 2, Application US/09735787  
Patent No. US20010036910A1  
GENERAL INFORMATION:  
APPLICANT: Rasmussen, Grethe  
Mikkelsen, Jan Koller  
Schulein, Martin  
Packer, Shankant A.  
Hagen, Fred  
TITLE OF INVENTION: A Cellulase Preparation Comprising an  
Endoglucanase Enzyme  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. US20010036910A10 No. US20010036910A1disk of No. US200100:  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/735,787  
FILING DATE: 13-Dec-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/189,028  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambitis, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3469,214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-735-787-2

## Query Match

Best Local Similarity 69.9%; Score 1159; DB 9; Length 305;  
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```

Qy 1 MSSTPVLTITLAAALPIVASASGSGSTRYWDCKPSCAMPKRAVSPVYACDANFOR 60
Db 1 MSSSLLPFAVVAALPIVLAAL--DGRSTRYWDCKPSCGMKAKAPVNOVPSCKANFOR 58
Qy 61 LSPFNVOGSCN-GGSAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 119
Db 59 ITDFNAKSGCEPGGVAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 118
Qy 120 PVAKTMYVOSTSTGDLGSGNOFDIAMEGGGVCIFNGCSGSPFGALPGAQYGISRRDCCD 179
Db 119 PVAKKMYVOSTSTGDLGSGNHFDLNIPEGGVCIFNGCTPFGALPGAQYGISRRNECD 178
Qy 180 SFPAALKPGCQRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDALKPGCYKRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDGNFPVAVQIPSSST 238
Qy 239 NGCTGPTSTAPSGQTS-----PGGSGGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLN 292
Db 239 SSPVNGPTSTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGKMSGCTTCVAGSTCTKIN 298
Qy 293 DYISGCL 299
Db 299 DWYHQCL 305

```

# RESULT 4

US-10-138-870-2  
 ; Sequence 2, Application US/10138870  
 ; Publication No. US20030119167A1

## GENERAL INFORMATION:

APPLICANT: Rasmussen, Gretche  
 Mikkelson, Jan Moller  
 Schulein, Martin  
 Packard, Shankant A.  
 Hagen, Fred  
 TITLE OF INVENTION: A Cellulase Preparation Comprising an  
 Endoglucanase Enzyme

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. US20030119167A10 No. US20030119167A1disk of No. US20030119167A1

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/138,870

FILING DATE: 03-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/735,787

FILING DATE: 13-Dec-2000

APPLICATION NUMBER: 09/189,028

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Landis, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3469,214-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 305 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

MSSTPVLTITLAAALPIVASASGSGSTRYWDCKPSCAMPKRAVSPVYACDANFOR 60  
 MSSSLLPFAVVAALPIVLAAL--DGRSTRYWDCKPSCGMKAKAPVNOVPSCKANFOR 58  
 LSPFNVOGSCN-GGSAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 119  
 ITDFNAKSGCEPGGVAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 118  
 PVAKTMYVOSTSTGDLGSGNOFDIAMEGGGVCIFNGCSGSPFGALPGAQYGISRRDCCD 179  
 PVAKKMYVOSTSTGDLGSGNHFDLNIPEGGVCIFNGCTPFGALPGAQYGISRRNECD 178  
 SFPAALKPGCQRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDSSFP-VFTPPSGG 238  
 RFPDALKPGCYKRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDGNFPVAVQIPSSST 238  
 NGCTGPTSTAPSGQTS-----PGGSGGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLN 292  
 SSPVNGPTSTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGKMSGCTTCVAGSTCTKIN 298  
 DYISGCL 299  
 DWYHQCL 305

# US-10-138-870-2

Query Match 69.9%; Score 1159; DB 14; Length 305;

Best Local Similarity 67.1%; Pred. No. 2,3e-89;

Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```

Qy 1 MSSTPVLTITLAAALPIVASASGSGSTRYWDCKPSCAMPKRAVSPVYACDANFOR 60
Db 1 MSSSLLPFAVVAALPIVLAAL--DGRSTRYWDCKPSCGMKAKAPVNOVPSCKANFOR 58
Qy 61 LSPFNVOGSCN-GGSAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 119
Db 59 ITDFNAKSGCEPGGVAYSCADQTPMAVNDMLAYGFAATSIAGSSSSWCCACVATLFTSG 118
Qy 120 PVAKTMYVOSTSTGDLGSGNOFDIAMEGGGVCIFNGCSGSPFGALPGAQYGISRRDCCD 179
Db 119 PVAKKMYVOSTSTGDLGSGNHFDLNIPEGGVCIFNGCTPFGALPGAQYGISRRNECD 178
Qy 180 SFPAALKPGCQRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDSSFP-VFTPPSGG 238
Db 179 RFPDALKPGCYKRFEDMFQANNDPTFTFOVQCPAEIYVARSCKRNDGNFPVAVQIPSSST 238
Qy 239 NGCTGPTSTAPSGQTS-----PGGSGGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLN 292
Db 239 SSPVNGPTSTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGKMSGCTTCVAGSTCTKIN 298
Qy 293 DYISGCL 299
Db 299 DWYHQCL 305

```

# RESULT 5

US-09-261-329-5  
 ; Sequence 5, Application US/09261329  
 ; Publication No. US20030092097A1

GENERAL INFORMATION:

APPLICANT: Andersen, Kim

APPLICANT: Schulein, Martin

APPLICANT: Christiansen, Lars

APPLICANT: Damgaard, Bo

APPLICANT: Von Der Osten, Claus

TITLE OF INVENTION: Cellulase Variants

FILE REFERENCE: 4887,204-US

CURRENT APPLICATION NUMBER: US/09/261,329

CURRENT FILING DATE: 1999-03-03

EARLIER APPLICATION NUMBER: 1013/96

EARLIER FILING DATE: 1996-09-17

NUMBER OF SEQ ID NOS: 26

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 201

TYPE: PRT

ORGANISM: Cellulase variants

US-09-261-329-5

Query Match 67.6%; Score 1122; DB 10; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-86;  
 Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 24 GSGGSTRYWDCKPSCAMPKRAVSPVYACDANFORLSDFNVOGSGNGSAYSCADQTP 83
Db 1 GSGGSTRYWDCKPSCAMPKRAVSPVYACDANFORLSDFNVOGSGNGSAYSCADQTP 60
Qy 84 NAYVNDMLAYGFAATSIAGSSSSWCCACVATLFTSGPVAGKMYVOSTSTGDLGSGNOFD 143
Db 61 NAYVNDMLAYGFAATSIAGSSSSWCCACVATLFTSGPVAGKMYVOSTSTGDLGSGNOFD 120
Qy 144 LMPGGGVGIFNGCSGSPFGALPGAQYGISRRDCCSFPAALKPGCQRFEDMFQANNDPT 203
Db 121 LMPGGGVGIFNGCSGSPFGALPGAQYGISRRDCCSFPAALKPGCQRFEDMFQANNDPT 180
Qy 204 PTFQVQCPAEIYVARSCKR 223

```

Db 181 FTQGVQCPAEIVASGCKR 200

RESULT 6  
US-08-841-636A-31  
Sequence 31, Application US/08841636A  
Publication No. US20020168751A1  
GENERAL INFORMATION:  
APPLICANT: Miettinen-Oinonen, Arja  
APPLICANT: Londeborough, John  
APPLICANT: Vehmanen, Jari  
APPLICANT: Haakana, Hei  
APPLICANT: Mityl, Arja  
APPLICANT: Lantto, Raija  
APPLICANT: Elvainen, Minna  
APPLICANT: Joutsen, Vesa  
APPLICANT: Palohimo, Maria  
APPLICANT: Suominen, Pirko  
TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND  
TITLE OF INVENTION: USES THEREOF  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/841,636A  
FILING DATE: 30-APR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,335  
FILING DATE: 17-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,926  
FILING DATE: 04-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,840  
FILING DATE: 28-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/732,181  
FILING DATE: 16-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: ECT/F196/00550  
FILING DATE: 17-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Timothy J. Shea, Jr.  
REGISTRATION NUMBER: 41,306  
REFERENCE/DOCKET NUMBER: 1716, 0510005/MAC/TJS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)371-2600  
TELEFAX: (202)371-2540  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Melanocarpus albomyces  
STRAIN: ALKO4237  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..235  
OTHER INFORMATION: /label= 20K-cellulase

US-08-841-636A-31

Query Match 62.6%; Score 1039; DB 8; Length 235;  
Best Local Similarity 76.3%; Pred. No. 2, 1e-79;  
Matches 180; Conservative 27; Mismatches 27; Indels 2; Gaps 1;

QY 1 NASTVLEKTLAALPLVVAASGSGSTYRWCKKSCAPGAAVSGPVYACCAACOR 60  
DB 1 KRSTVLRALIAALPLGALAA--NGSTYRWCKKSCGWRGKGPVNGPVYSCANRFR 58  
QY 61 LSDENWGCGGSAVSCADOTPAVNDNAYFAATSIAGSSSSMCCCYALTFTSGP 120  
DB 59 IHPDPAVSGGPGAPSCADHSFPAINDNLSYGFATLISGTEBSMCCCYALTFTSGP 118  
QY 121 VAGKTMVVGSTGCDLGSNCPDIAIMPQGVGIENGCSQSFGSLPGAQYGGISSRDQDS 180  
DB 119 VAGKTMVVGSTGCDLGSNCPDIAIMPQGVGIENGCSQSFGSLPGAQYGGISSRDQDS 178  
QY 181 FPAELKPGCCWRPWFQNDNPTFTFOQVCCPAEIVASGCKRNDSSFVFTPPS 236  
DB 179 FPEPLKPGCCWRPWFQNDNPTFTFERVCCPELVAKTGCKRRHDDGGFAVFAAPS 234

RESULT 7

US-10-007-521-6

Sequence 6, Application US/10007521  
Publication No. US20030054539A1  
GENERAL INFORMATION:  
APPLICANT: Schuelein, Martin  
APPLICANT: Andersen, Lene N.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Lange, Lene  
APPLICANT: Nielsen, Ruby I.  
APPLICANT: Ihara, Michiko  
APPLICANT: Takagi, Shinobu  
TITLE OF INVENTION: No. US20030054539A1el Endoglucanases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. US20030054539A1o No. US20030054539A1disk of No. US200300  
STREET: 405 Lexington Avenue, 64th floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/007,521  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136  
FILING DATE: 21-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366, 200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 308 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-007-521-6

Query Match: 65.5%; Score 1021; DB 14; Length 308;  
Best Local Similarity 60.9%; Pred. No. 9.4e-78;  
Matches 185; Conservative 43; Mismatches 62; Indels 14; Gaps 7;

QY 9 TTAALPLVA-SAASGSGSTRYWDCCPSGKAMPKAAVSQVYACANFORLSD-FNV 66  
DB 6 TTGFLALPVALDQLSGIGQTRRYWDCCPSGKAMPKGP-SSPVQACDKNUNPLMDGSGT 64  
QY 67 QGCGN-GGSAVSCADDTPMANVNDNLATGPAATSIAGSSBSWCCACTALTFTSGEVAKT 125  
DB 65 RSGCGAGSAYMCSQSPWAVSDLSYGAAYVLAGSSBSQWCCACCELTFTSGEVAKK 124  
QY 126 MVVOSTSTGDLGSDNFDIAMPGGSGVGFNGCSSQFGALP---GAQYGGISRDQDSFP 182  
DB 125 MIVQATNTGDLGDNHFDLALPGGSGVGFNACTDQYGAPEMGWDRYGGIHSKECESFP 184  
QY 183 APLKPGCQMRPFQNDNPTFTFQVQCPAEIVASGCKNDSSFP-VTPPSGANGG 241  
DB 185 EALKPGCQMRPFQNDNPTFTFQVQCPAEIVASGCKNDSSFP-VTPPSGANGG 244  
QY 242 TGTPTSTARGSGQTS-----PGGSGCTSQKAAQCGSIGSGCTTCVSGTTCQKLNXY 295  
DB 245 VNQPTSTSTSTSTSSPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKINDMY 304  
QY 296 SQCL 299  
DB 305 HQCL 308

RESULT 8  
US-10-007-521-4

? Sequence 4, Application US/10007521  
? Publication No. US20030054539A1  
? GENERAL INFORMATION:  
? APPLICANT: Schulein, Martin  
? Andersen, Lene N.  
? Lassen, Soren F.  
? Kauppinen, Markus S.  
? Lange, Lene  
? Nielsen, Ruby I.  
? Ihara, Michiko  
? Takagi, Shinobu  
? TITLE OF INVENTION: No. US20030054539A1el Endoglucanases  
? NUMBER OF SEQUENCES: 109  
? CORRESPONDENCE ADDRESS:  
? ADDRESS: No. US20030054539A1o No. US20030054539A1disk of No. US20030054539A1  
? STREET: 405 Lexington Avenue, 64th Floor  
? CITY: New York  
? STATE: New York  
? COUNTRY: United States of America  
? ZIP: 10174-6401  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patentin Release #1.0, Version #1.30  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/10/007,521  
? FILING DATE: 10-Dec-2001  
? CLASSIFICATION: <Unknown>  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/08/651,136  
? FILING DATE: 21-MAY-1996  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Lambiris, Elias J.  
? REGISTRATION NUMBER: 33,728  
? REFERENCE/DOCKET NUMBER: 4366.200-US  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 212-867-0123  
? TELEFAX: 212-878-9655  
? INFORMATION FOR SEQ ID NO: 4:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 297 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-007-521-4

Query Match: 59.9%; Score 994.5; DB 14; Length 297;  
Best Local Similarity 60.7%; Pred. No. 1.5e-75;  
Matches 181; Conservative 45; Mismatches 59; Indels 13; Gaps 8;

QY 9 TTAALPLVA-SAASGSGSTRYWDCCPSGKAMPKAAVSQVYACANFORLSD-FNV 66  
DB 6 TTGFLALPVALDQLSGIGQTRRYWDCCPSGKAMPKGP-SSPVQACDKNUNPLMDGSGT 64  
QY 67 QGCGN-GGSAVSCADDTPMANVNDNLATGPAATSIAGSSBSWCCACTALTFTSGEVAKT 125  
DB 65 RSGCGAGSAYMCSQSPWAVSDLSYGAAYVLAGSSBSQWCCACCELTFTSGEVAKK 124  
QY 126 MVVOSTSTGDLGSDNFDIAMPGGSGVGFNGCSSQFGALP---GAQYGGISRDQDSFP 182  
DB 125 MIVQATNTGDLGDNHFDLALPGGSGVGFNACTDQYGAPEMGWDRYGGIHSKECESFP 184  
QY 183 APLKPGCQMRPFQNDNPTFTFQVQCPAEIVASGCKR-NDSSFPVTPPSGANGG 241  
DB 185 EALKPGCQMRPFQNDNPTFTFQVQCPAEIVASGCKR-NDSSFPVTPPSGANGG 242  
QY 242 TGTPTSTARGSGQTS-----PGGSGCTSQKAAQCGSIGSGCTTCVSGTTCQKLNXY 295  
DB 245 VNQPTSTSTSTSTSSPVQPTTPSGCTAERMAQCGGNGSGCTTCVAGSTCTKINDMY 304

RESULT 9  
US-10-007-521-22

? Sequence 22, Application US/10007521  
? Publication No. US20030054539A1  
? GENERAL INFORMATION:  
? APPLICANT: Schulein, Martin  
? Andersen, Lene N.  
? Lassen, Soren F.  
? Kauppinen, Markus S.  
? Lange, Lene  
? Nielsen, Ruby I.  
? Ihara, Michiko  
? Takagi, Shinobu  
? TITLE OF INVENTION: No. US20030054539A1el Endoglucanases  
? NUMBER OF SEQUENCES: 109  
? CORRESPONDENCE ADDRESS:  
? ADDRESS: No. US20030054539A1o No. US20030054539A1disk of No. US20030054539A1  
? STREET: 405 Lexington Avenue, 64th Floor  
? CITY: New York  
? STATE: New York  
? COUNTRY: United States of America  
? ZIP: 10174-6401  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patentin Release #1.0, Version #1.30  
? CURRENT APPLICATION DATA:  
? APPLICATION NUMBER: US/10/007,521  
? FILING DATE: 10-Dec-2001  
? CLASSIFICATION: <Unknown>  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: US/08/651,136  
? FILING DATE: 21-MAY-1996  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Lambiris, Elias J.  
? REGISTRATION NUMBER: 33,728  
? REFERENCE/DOCKET NUMBER: 4366.200-US  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 212-867-0123  
? TELEFAX: 212-878-9655  
? INFORMATION FOR SEQ ID NO: 22:

## SEQUENCE CHARACTERISTICS:

LENGTH: 310 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 22  
US-10-007-521-22

Query Match 58.0%; Score 962; DB 14; Length 310;  
Best Local Similarity 55.9%; Pred. No. 8.7e-73;  
Matches 175; Conservative 46; Mismatches 72; Indels 20; Gaps 5;

QY 1 MESTPVLRTTAAALPLVSAASGSGSTRYMDCKPCSCAMPGRGAASQPVYACDANFOR 60  
1 MRSAAVLIGLVAG---VAAQSSGSTRITRYMDCKPCSCAMPGRGAASQPVYACDANFOR 56  
DB 61 LSPFNVSQSC-NGSAYSCADQTPMAYNDNLAYGPAATSIAGSSSSWCCACTALFTSG 119  
57 LAS-TARSCDSNVAATYCNQDNPAYVNDNLAYGPAATSIAGSSSSWCCACTALFTSG 115  
QY 120 PVAGKTVVQSTSTGDLGSDNPDLPAGGVGIFNGCSSQFG-GLPQAQVGISSRDQC 178  
DB 116 PVAGKTVVQSTSTGDLGSDNPDLPAGGVGIFNGCSSQFG-GLPQAQVGISSRDQC 175  
QY 179 DSPFAPLPKPCQMRPFQNDNPTFTFOQVCPAETIYASGCKRNDSSFPVTPP--- 235  
DB 176 SLPKALQPCQMRPFQNDNPTFTFOQVCPAETIYASGCKRNDSSFPVTPP--- 235  
QY 236 ATTIRITTTTASSSSSSSTTASGSPVFTGGSGGPTSPVWQCGGQGMSGPTSCVAG 295  
QY 286 TTCOKLNDYSSQC 298  
DB 236 STCSVNPATYSQC 308

## RESULT 10

US-10-007-521-8  
Sequence 8, Application US/10007521  
Publication No. US20030054539A1  
GENERAL INFORMATION:

APPLICANT: Schulein, Martin  
Andersen, Lene N.  
Lassen, Soren F.  
Kauppinen, Markus S.  
Lange, Lene  
Nielsen, Rudy I.  
Ihara, Michiko  
Takagi, Shinobu  
TITLE OF INVENTION: No. US20030054539A1 Endoglucanases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: No. US20030054539A1 No. US20030054539A1disk of No. US20030054539A1  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/007,521  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/651,136  
FILING DATE: 21-May-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366,200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 295 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-10-007-521-8

Query Match 57.3%; Score 951; DB 14; Length 295;  
Best Local Similarity 58.5%; Pred. No. 6.9e-72;  
Matches 166; Conservative 37; Mismatches 78; Indels 10; Gaps 5;

QY 1 MESTPVLRTTAAALPLVSAASGSGSTRYMDCKPCSCAMPGRGAASQPVYACDANFOR 60  
1 MRSIILIGLVAG---VAAQSSGSGHTTRYMDCKPCSCAMPGRGAASQPVYACDANFOR 56  
DB 61 LSPFNVSQSC-NGSAYSCADQTPMAYNDNLAYGPAATSIAGSSSSWCCACTALFTSG 119  
QY 57 LSPGAV-SGCDPVGVAFTCNQDNPAYVNDNLAYGPAATSIAGSSSSWCCACTALFTSG 115  
DB 120 PVAGKTVVQSTSTGDLGSDNPDLPAGGVGIFNGCSSQFG-GLPQAQVGISSRDQC 178  
QY 116 PVAGKTVVQSTSTGDLGSDNPDLPAGGVGIFNGCSSQFG-GLPQAQVGISSRDQC 175  
DB 179 DSPFAPLPKPCQMRPFQNDNPTFTFOQVCPAETIYASGCKRNDSSFPVTPP--- 238  
QY 176 AELPSVLAQPCQMRPFQNDNPTFTFOQVCPAETIYASGCKRNDSSFPVTPP--- 232  
DB 239 NGCTGTPSTIAPSGGTSIPGSSGCTSQKAAQCGGIGTTCVSGTTTCOKLNDYSSQC 298  
QY 233 TGPPTPTTITTSPPQPTKGGGCTSPHWQCGGQGMSGPTSCVAG 292  
QY 299 L 299  
DB 293 I 293

## RESULT 11

US-09-735-787-4  
Sequence 4, Application US/09735787  
Patent No. US20010036910A1  
GENERAL INFORMATION:

APPLICANT: Rasmussen, Grethe  
Mikkelsen, Jan Moller  
Schulein, Martin  
Packar, Shankar A.  
Hagen, Fred  
TITLE OF INVENTION: A Cellulase Preparation Comprising an  
Endoglucanase Enzyme  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: No. US20010036910A1 No. US20010036910A1disk of No. US20010036910A1  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/735,787  
FILING DATE: 13-Dec-2000  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/189,028  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3469,214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-735-787-4

Query Match 56.6%; Score 939.5; DB 9; Length 376;  
Best Local Similarity 49.5%; Pred. No. 8,5e-71;  
Matches 183; Conservative 35; Mismatches 73; Indels 79; Gaps 6;

QY 9 TTTAALPLVSAASGSGSTRYDCCXPCSCAPKAASQPVYACDANFQRLSDPFWQS 68  
DB 5 TTTAAGPLAVSAASGSGSTRYDCCXPCSCAPKAASQPVYACDANFQRLSDPFWQS 64  
QY 69 GC-NGSAYSCADQTPMAVNDLAVGPAATSIAGSSESWCCACVALTFTSGPVAGKTV 127  
DB 65 GEGGGSAYACTNYSPMAVNDLAVGPAATSIAGSSESWCCACVALTFTSGPVAGKTV 124  
QY 128 VQSTGTGDLGSDNPDIAMPGGVGIENGCSQPG-GLPGAQYGGISSSDQCDSPAPLX 186  
DB 125 VQSTGTGDLGSDNPDIAMPGGVGIENGCSQPG-GLPGAQYGGISSSDQCDSPAPLX 184  
QY 187 PCQWRPFQADNPFTFQOVCPAEIVASGCKRNDSSFPVF----- 232  
DB 185 DQCHRFDFENADNPFTFQOVCPKALIDISGCKRNDSSFPKAVDTASKRPQSS 244  
QY 233 -----TPPS-----GNGGTPPTSTAPSGQTSF 257  
DB 245 AKTTSAALAAQPKTKDSAPVQKSTKPAQPEPTKPADKQTDKPVATKPAATKPVQ 304  
QY 258 GGGSGCTSQK-----WAQCGIGF---SGCTTCVSGTTCC 289  
DB 305 PUNKKTTQKVGKTRGSCPAKTDATKASVVPAYVCGGSKSAYPNGMLACATGSKCV 364  
QY 290 KLDYYSQCL 299  
DB 365 KQNEYYSQCV 374

RESULT 12  
US-10-138-870-4  
Sequence 4, Application US/1013870  
Publication No. US20030119167A1  
GENERAL INFORMATION:  
APPLICANT: Rasmussen, Gretche  
Mikkelsen, Jan Moller  
Schulein, Martin  
Patkar, Shankant A.  
Hagen, Fred  
TITLE OF INVENTION: A Cellulase Preparation Comprising an  
Endoglucanase Enzyme  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. US20030119167A; No. US20030119167A; disk of No. US20030119167A  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDICAL TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/138,870  
FILING DATE: 03-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/735,787  
FILING DATE: 13-Dec-2000  
APPLICATION NUMBER: 09/189,028  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3469,214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-138-870-4

Query Match 56.6%; Score 939.5; DB 14; Length 376;  
Best Local Similarity 49.5%; Pred. No. 8,5e-71;  
Matches 183; Conservative 35; Mismatches 73; Indels 79; Gaps 6;

QY 9 TTTAALPLVSAASGSGSTRYDCCXPCSCAPKAASQPVYACDANFQRLSDPFWQS 68  
DB 5 TTTAAGPLAVSAASGSGSTRYDCCXPCSCAPKAASQPVYACDANFQRLSDPFWQS 64  
QY 69 GC-NGSAYSCADQTPMAVNDLAVGPAATSIAGSSESWCCACVALTFTSGPVAGKTV 127  
DB 65 GEGGGSAYACTNYSPMAVNDLAVGPAATSIAGSSESWCCACVALTFTSGPVAGKTV 124  
QY 128 VQSTGTGDLGSDNPDIAMPGGVGIENGCSQPG-GLPGAQYGGISSSDQCDSPAPLX 186  
DB 125 VQSTGTGDLGSDNPDIAMPGGVGIENGCSQPG-GLPGAQYGGISSSDQCDSPAPLX 184  
QY 187 PCQWRPFQADNPFTFQOVCPAEIVASGCKRNDSSFPVF----- 232  
DB 185 DQCHRFDFENADNPFTFQOVCPKALIDISGCKRNDSSFPKAVDTASKRPQSS 244  
QY 233 -----TPPS-----GNGGTPPTSTAPSGQTSF 257  
DB 245 AKTTSAALAAQPKTKDSAPVQKSTKPAQPEPTKPADKQTDKPVATKPAATKPVQ 304  
QY 258 GGGSGCTSQK-----WAQCGIGF---SGCTTCVSGTTCC 289  
DB 305 PUNKKTTQKVGKTRGSCPAKTDATKASVVPAYVCGGSKSAYPNGMLACATGSKCV 364  
QY 290 KLDYYSQCL 299  
DB 365 KQNEYYSQCV 374

RESULT 13  
US-09-261-329-4  
Sequence 4, Application US/09261329  
Publication No. US20030092097A1  
GENERAL INFORMATION:  
APPLICANT: Andersen, Kim  
Schulein, Martin  
APPLICANT: Christiansen, Lars  
APPLICANT: Damgaard, Bo  
APPLICANT: Von der Oetel, Claus  
TITLE OF INVENTION: Cellulase Variants  
FILE REFERENCE: 4887,204-US

CURRENT APPLICATION NUMBER: US/09/261,329  
CURRENT FILING DATE: 1999-03-03  
EARLIER APPLICATION NUMBER: 1013/96  
EARLIER FILING DATE: 1996-09-17  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 201  
TYPE: PRC  
ORGANISM: Cellulase variants  
US-09-261-329-4

Query Match 56.5%; Score 938; DB 10; Length 201;  
Best Local Similarity 79.0%; Pred. No. 5,66-71;  
Matches 158; Conservative 26; Mismatches 16; Indels 0; Gaps 0;

QY 24 GSGQSTRYWDCKPCSCAMPKRAVSPYVACDANFQRLSDPENVQSGNGGSAVSCADQTP 83  
1 GSGKSTRYWDCKPCSCAMPKRAVSPYVACDANFQRLSDPENVQSGNGGSAVSCADQTP 60  
QY 84 WANDLAFPAATSIAGSESSMCCACALFTTSPGVAKTMVOSTSTGDLGSHQPD 143  
61 WANDLAFPAATSIAGSESSMCCACALFTTSPGVAKTMVOSTSTGDLGSHQPD 120  
QY 144 IAMPGGVGIENGCSQFQGLPGAQYGGISRDQDSFPAPLKPCCQMRPFWQNDNPT 203  
121 IAMPGGVGIENGCSQFQGLPGAQYGGISRDQDSFPAPLKPCCQMRPFWQNDNPT 180  
QY 204 PTFQOVQCPAETVARSGCKR 223  
181 PTFQOVQCPAETVARSGCKR 200  
DB

RESULT 14  
US-10-007-521-24  
Sequence 24, Application US/10007521  
Publication No. US20030054539A1  
GENERAL INFORMATION:

APPLICANT: Schulein, Martin  
Andersen, Lene N.  
Lassen, Soren F.  
Kauppinen, Markus S.  
Lange, Lene  
Nielsen, Ruby I.  
Ihara, Michiko  
Takagi, Shinobu  
TITLE OF INVENTION: No. US20030054539A1 Endogluconases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSER: No. US20030054539A10 No. US20030054539A1disk of No. US20030054539A1  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/007,521  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136  
FILING DATE: 21-MAY-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 294 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-10-007-521-24

Query Match 55.7%; Score 923.5; DB 14; Length 294;  
Best Local Similarity 57.9%; Pred. No. 1,46-69;  
Matches 169; Conservative 43; Mismatches 69; Indels 11; Gaps 6;

QY 13 MALPVSASAGSGQSTRYWDCKPCSCAMPKRAVSPYVACDA--NFQRLSDPENVQSGC 70  
9 MALPVSASAGSGQSTRYWDCKPCSCAMPKRAVSPYVACDA--NFQRLSDPENVQSGC 67  
QY 71 NGGSAYSCADQTPWANDLAFPAATSIAGSESSMCCACALFTTSPGVAKTMVOST 130  
68 NGGSAYSCADQTPWANDLAFPAATSIAGSESSMCCACALFTTSPGVAKTMVOST 127  
QY 131 TSTGDLGSHQPD IAMPGGVGIENGCSQFQGLPGAQYGGISRDQDSFPAPLKPCC 183  
128 TSTGDLGSHQPD IAMPGGVGIENGCSQFQGLPGAQYGGISRDQDSFPAPLKPCC 187  
QY 190 QMRPFWQNDNPT PTFQOVQCPAETVARSGCKR 223--SFPVTFPSSGNGGTGTPTS 247  
188 QMRPFWQNDNPT PTFQOVQCPAETVARSGCKR 223--SFPVTFPSSGNGGTGTPTS 245  
QY 248 TAPSGQTSFGSGGCTGSGXMAQCGGIGFSGCTTCVSGTTCQKINDYSGCT 299  
246 TAPSGQTSFGSGGCTGSGXMAQCGGIGFSGCTTCVSGTTCQKINDYSGCT 294  
DB

RESULT 15  
US-10-007-521-10  
Sequence 10, Application US/10007521  
Publication No. US20030054539A1  
GENERAL INFORMATION:

APPLICANT: Schulein, Martin  
Andersen, Lene N.  
Lassen, Soren F.  
Kauppinen, Markus S.  
Lange, Lene  
Nielsen, Ruby I.  
Ihara, Michiko  
Takagi, Shinobu  
TITLE OF INVENTION: No. US20030054539A1 Endogluconases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSER: No. US20030054539A10 No. US20030054539A1disk of No. US20030054539A1  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/007,521  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136  
FILING DATE: 21-MAY-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366.200-US

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 349 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-007-521-10

Query Match 54.9%; Score 910.5; DR 14; Length 349;

Best Local Similarity 50.1%; Pred. No. 2.1e-68;

Matches 168; Conservative 41; Mismatches 77; Indels 49; Gaps 6;

```
QY 11 LAALPLVASAASGSGSTRWYDCCPKPCAMPKAAVSQPVYACDANFORLSDPNVQSQC 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 LVGAVPSSVMAASGKHTRWYDCCCKTSCKMEGKASVSEPVLTCKQDNPIVDANARSGC 69
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 71 NGSAYSCADQTPAAVNDNLAYFPAISIAQSSSSWCCACVALFTSGFPVAGKTWVQS 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 DGGAFACITNSPVAIVEDLAYFAATLSCGTGSMCCACVALFTSGFPVAGKTWVQS 129
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 TSTGELISNOFDIAMPGGVGITENGCSORFG-LPGAQYGISRPDCDSFPAPLKEQC 189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 130 TATGDLISNNHFDLIPGGGLGIFGSCSPGQLPGERYGVSRSQCIDMPELHKDQC 189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 190 QWRFDPMQALNPTFTFOVQCPAELIVARSGCKRNDSSFPVETTPSGGNGGTGTP--- 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 190 QWRFDPMQALNPTFTFOVQCPAELIVARSGCKRNDSSFPVETTPSGGNGGTGTP--- 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 -----ISTAPG-----SGQISPGG-----GSQCT 264
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 248 TTTSSEKPTISAPSTISNPSPAPQPGWTDPAETTTTKLPALPATTSPPAVSVPSRAR 307
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 SQKAAQCGG-IGFGCTTCVSGTTCQKLDYVSQC 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 VPLMGQCDSEASWDAEKKCAKGTKCVYVNDYVSQC 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 7, 2004, 08:47:50  
Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using SW model

Run on: July 7, 2004, 08:48:30 ; Search time 21 Seconds  
(without alignments)  
1369,584 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSIPVLRITAAALPLVAS.....TTCVSGTTCCKNDYISQCL 299

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.78:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	740.5	44.6	229	2 JC7308	cellulase (EC 3.2.1.4)
2	507	30.6	511	2 S10527	endoglucanase B pr
3	462	27.8	393	2 S59499	cellulase egli - s
4	221	13.3	242	2 S60143	cellulase (EC 3.2.1.4)
5	219.5	13.2	513	1 EUNOI	cellulose 1,4-beta
6	219.5	13.2	513	2 S45380	cellulose 1,4-beta
7	215.5	13.0	513	2 S11439	cellulase 1,4-beta
8	214	12.9	459	2 A25928	cellulase (EC 3.2.1.4)
9	203	12.2	463	2 A48375	cellulase (EC 3.2.1.4)
10	190	11.5	516	2 S33164	cellulase 1,4-beta
11	189	11.4	464	2 JC7143	cellulase I -
12	188.5	11.4	302	2 S71334	acetyl xylan ester
13	185	11.2	525	1 S38794	cellulase 1,4-beta
14	184	11.1	516	2 JS0083	cellulase 1,4-beta
15	183	11.0	537	1 JU0150	cellulase 1,4-beta
16	181	10.9	516	2 S42093	cellulase 1,4-beta
17	172	10.3	320	2 JC1311	cell protein precu
18	170.5	10.3	510	2 S41943	cellulase 1,4-beta
19	170.5	10.3	511	2 S44716	cellulase 1,4-beta
20	161	9.7	504	1 B48939	cellulase 1,4-beta
21	156	9.4	856	2 T00349	Avicelase 111 - As
22	150.5	9.1	860	2 S43846	xylanase 3 - rumen
23	143.5	8.6	388	1 JC5461	cellulase (EC 3.2.1.4)
24	143.5	8.6	388	1 S43920	cellulase (EC 3.2.1.4)
25	141.5	8.5	530	2 JC7979	cellulohydrolase
26	139.5	8.4	540	2 S41942	cellulase 1,4-beta
27	137	8.3	438	1 S70602	cellulase 1,4-beta
28	137	8.3	839	2 F75518	hypothetical prote
29	129	7.8	418	1 S28372	cellulase (EC 3.2.1.4)

30	127	7.7	429	1 JC5861	endo-1,4-beta-xyla
31	125.5	7.6	316	1 A38743	loricin - human
32	125.5	7.6	410	1 S68153	cellulase (EC 3.2.1.4)
33	124	7.5	471	1 A26160	cellulose 1,4-beta
34	124	7.5	471	1 A38979	cellulose 1,4-beta
35	122	7.4	182	2 A36686	ultra-high-sulfur
36	120.5	7.3	303	2 H70716	hypothetical prote
37	120	7.2	5376	2 T42215	vestigial protein
38	118.5	7.1	453	2 A41640	triphosphin - human
39	118.5	7.1	749	2 I38488	cellobiohydrolase
40	117	7.1	459	2 JC7931	hypothetical glyci
41	116.5	7.0	584	2 G70804	keratin KAP5.4 - s
42	115.5	7.0	191	2 T46412	related to spore c
43	114	6.9	461	2 T51044	hypothetical prote
44	114	6.9	1804	2 H96597	hypothetical glyci
45	113.5	6.8	1381	2 E7806	hypothetical glyci

#### ALIGNMENTS

##### RESULT 1

JC7308  
cellulase (EC 3.2.1.4) - Scopulariopsis brevicaulis  
N:Alternate names: endoglucanase I  
C:Species: Scopulariopsis brevicaulis  
C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: JC7308; EC7087  
C:Nakatan, F.; Kawaguchi, T.; Takada, G.; Sumitani, J.; Moriyama, Y.; Arai, M.  
Biosci. Biotechnol. Biochem. 64, 1338-1346, 2000  
A:Title: Cloning and sequencing of an endoglucanase gene from Scopulariopsis brevicaulis  
A:Reference number: JC7308  
A:Accession: JC7308  
A:Molecule type: DNA  
A:Residues: 1-229 <NAK>  
A:Experimental source: strain TOF-12:2  
A:Accession: PC7087  
A:Molecule type: Protein  
A:Residues: 21-37,149-164 <NA2>  
A:Gene: egl  
A:Introns: 147/3  
C:Keywords: glycosidase; hydrolase

Query Match 44.6%; Score 740.5; DB 2; Length 229;  
Best Local Similarity 59.8%; Pred. No. 2.6e-45;  
Matches 134; Conservative 32; Mismatches 49; Indels 9; Gaps 5;

QY	10	TLAAALPLV---ASNAAGSGQSTRWDCCKPSCAMPGRRAVSQ-PYVACDANFOLSDFN 65
DB	6	TLALTLFLVLPASQASGSTRTRWDCCKPSCMPDRAPLSQGPWTCIDINDNPLDDG 65
QY	66	V-GSGCC-GGASVACADOTPMANVDNLAQFAATSIAGSSSSSWCCACVLAFTTSGPAG 123
DB	66	LTESGCPGGGAGACSSHSFPAVDDDELAYGMAVNIQQTSSDWCCACVLEFTTGAVS 125
QY	124	KTMVVGSTGSDIGSNGEDLAMPGGVGLPFGCSSQPGGLP---GAQYGISSDQDS 180
DB	124	KMTVVGKTMVVGSDIGSNGEDLAMPGGVGLPFGCSSQPGGLP---GAQYGISSDQDS 185
QY	181	FPAPLKPCQWRPDMFONADNPFTFGQVQCPAETIARSSCKRN 224
DB	186	FPEALKAGCEMRDWFQGDNDPVSFREVCEPAETIARSSQCKRS 229

##### RESULT 2

S10527  
endoglucanase B precursor - Pseudomonas fluorescens  
C:Species: Pseudomonas fluorescens  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
C:Accession: S10527  
C:Gilbert, H.J.; Hall, U.; Hazlewood, G.P.; Ferreira, L.M.A.  
Mol. Microbiol. 4, 759-767, 1990

A>Title: The N-terminal region of an endoglucanase from *Pseudomonas fluorescens* subsp. *A*  
A:Reference number: S10527; MUID:9035856; PMID:2117653  
A:Accession: S10527  
A:Status: pre-iterinary  
A:Molecule type: DNA  
A:Residues: 1-511 <GIL>  
A:Cross-references: EMBL:X52615; NID:945497; PIRN:CA16844.1; PID:945498  
C:Superfamily: bacterial cellulose-binding domain homology; glycosylase GWC domain hom  
F:31-128/Domain: bacterial cellulose-binding domain homology <BCB>  
F:180-217/Domain: glycosylase GWC domain homology <GWC>  
F:32-127/Disulfide bonds: #status predicted

Query Match 30.6%; Score 507; DB 2; Length 511;  
Best Local Similarity 39.6%; Pred. No. 1.5e-28;

Matches 107; Conservative 38; Mismatches 73; Indels 52; Gaps 7;

QY 2 RSTPVLRTTAAALPVVSAASG-----SGQSTRYMDCKPSCAPGKA-AVSPQY 52  
DB 237 ASSVASSSSLSATSSASVSPETDGGNGVATRYMDCKPSCGMSAVPGLVPIQ 296  
QY 53 ACDANFRLSDPNVSGGSAVSCADQTPAVVNDLAVGPAATSIAGSSSSWCCATC 112  
DB 297 SCGANWTRLDVSVSSCDGGGCMCDIPAVSPITAVGAAAT-----SSGVCGRCY 351  
QY 113 ALFTT-----SGPVAGTMYVOSTSTGDLGANOFDIAMPGGGVIPIFGSSQF 161  
DB 352 QLOFTSSVNAPEDEGSAALAKMTIVQNTNIGDVSGGFDLIVGGGVAHAPACAG 411  
QY 162 G--GFGAYGIGISSR-----DQCDSF-----PAPKPGCQWRD 194  
DB 412 GVSNAEAGAYGGFLAAKQGLAGYNASTCYKSCVLRCDSPVSGRLTQLQGQCTWFA 471  
QY 195 WFOANDPPTPTQVQCPAETIVASGCKN 224  
DB 472 WFEADNPISLTKETPCPAELTRSGMRS 501

## RESULT 3

S59499  
cellulase eg1 - smut fungus (Ustilago maydis)  
C:Species: Ustilago maydis (corn smut)  
C:Date: 20-Jun-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999  
C:Accession: S59499  
R:Schauwecker, F.; Wanner, G.; Kahmann, R.  
Bio1. Chem. Hoppe-Seyler 376, 617-625, 1995  
A>Title: Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago*  
A:Reference number: S59499; MUID:96145728; PMID:8590631  
A:Accession: S59499  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-393 <SCH>  
A:Cross-references: GB:S81598; NID:91478378; PIRN:AAB36147.1; PID:91478379

Query Match 27.8%; Score 462; DB 2; Length 393;  
Best Local Similarity 30.9%; Pred. No. 1.8e-25;

Matches 115; Conservative 41; Mismatches 100; Indels 116; Gaps 12;

QY 11 LAALPLVVSASGSGSTRYMDCKPSCAPGKAASQPTVACDANPRL-----SDPY 66  
DB 12 LSLSLVHLIDGVRAMATRYMDCLASMSGKAPVAPYADCKADGVTLLIDSKKPSG 71  
QY 67 QSGNGSASVSCADQTPW--AVNDNLAVGPATSIAGSSSSWCCACVAFPTSPVAK 124  
DB 72 QSGNGSASVSCADQTPW--AVNDNLAVGPATSIAGSSSSWCCACVAFPTSPVAK 124  
QY 125 TW-----VWGSTSTGDLGANOFDIAMPGGGVIPIFGSSQF-----GLPGAQYIGISSR 175  
DB 129 AMKRNKILFQVTVAGGVQSONFDFQIPGGGAGAPKCPAQGVASIMGDOYGVGYSA 188  
QY 176 DQCSFPAPKSGCCQWRP--WFOANDPPT--TPOGVQCPAETIVASGCKNDSPPVE 232  
DB 189 TEGSKLPKPLQECCKRSEW---GDNFVAKSGPVKACPSLIDSGCORQDNTTISPY 245

QY 233 T----- 233  
DB 246 SCKDVSANTAAPQYKRBRSVCLAGKKGSAAGVDSGDASGADASGAGAAESG 305  
QY 234 -----PPSGGN-----GGTGTPTSTAPSGQTSPPGQ 260  
DB 306 QPEGVQPPSGGNQGSNNDATGTAGSGSGSDSGTANGSGGAPTSGSASVAPPGG 365  
QY 261 SGGTSQKMAQCG 272  
DB 366 SNPGAQGGQGG 377

## RESULT 4

S60143  
cellulase (EC 3.2.1.4) V precursor - fungus (Trichoderma reesei) (strain QM9414)  
N:Alternate names: endo-1,4-beta-glucanase V  
C:Species: Trichoderma reesei  
A:Variety: strain QM9414  
C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 17-Mar-1999  
A:Accession: S60143; S49043  
R:Salohimo, A.; Henriksen, B.; Hoffren, A.; Telemann, O.; Penttila, M.  
submitted to the EMBL Data Library, May 1994  
A:Description: A novel small endoglucanase gene eg15 from *Trichoderma reesei* isolated  
A:Reference number: S60143  
A:Accession: S60143  
A:Molecule type: DNA  
A:Residues: 1-242 <SNL>  
A:Cross-references: EMBL:Z33381; NID:9485863; PID:9485864  
R:Salohimo, A.; Henriksen, B.; Hoffren, A.M.; Telemann, O.; Penttila, M.  
Mol. Microbiol. 13, 219-228, 1994  
A>Title: A novel, small endoglucanase gene, eg15, from *Trichoderma reesei* isolated by  
A:Reference number: S49043; MUID:95075308; PMID:7984103  
A:Accession: S49043  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 206-241 <SNM>  
A:Cross-references: EMBL:Z33381  
C:Genetics:  
A:Gene: eg15  
A:Insertions: 23/3; 45/3  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as  
A:Pathway: cellulose degradation  
C:Superfamily: fungal cellulose-binding domain homology  
C:Keywords: glycosylase, hydrolase, polysaccharide degradation  
F:1-17/Domain: signal sequence #status predicted <Sig>  
F:18-242/Product: cellulase #status predicted <Mat>  
F:210-241/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.3%; Score 221; DB 2; Length 242;  
Best Local Similarity 24.7%; Pred. No. 1.1e-08;

Matches 78; Conservative 39; Mismatches 101; Indels 98; Gaps 13;

QY 7 LRTTLAALPLVVSASGSGSTRYMD-----CKPSCAPGKAASQPTVACDANPQ 59  
DB 1 MKATLVGLSLVAGVSAVYKATTRYDQEGACGSSSGAPFWGLIGMAYTA--AGSQ 59  
QY 60 RLSEFVNSGNGSASVSCADQTPAVVNDLAVGPAATSIAGSSSSWCCACVAFPT-- 117  
DB 60 ALFDTAGASWCGAG-----CGKCYQLTSTQ 85  
QY 118 -----SGPVAGTMYVOSTSTGDLGANOFDIAMPGGGVIPIFGSSQFPGALPAQY 170  
DB 86 APSSGSGTGAAGSILITVWTLCPNNGNMQCPV--GGTN-----QY-- 127  
QY 171 GISSRDQDSFPADLPKQCMRPD--WFOANDPPT--ADNPTFTFQVQCPAETIVASG--CKR 223  
DB 128 -----GYSYHFDIAGNEIFGDNVAVFEBIACPGGASDWGTCLCVG 170  
QY 224 NDDSPVFTPPSGNGGTGTPTSTAPSGQTSPPGSGSCTSQKMAQCGGIGSGCTTCV 283  
DB 171 QQRTD-P--TIVLNDTGSTIPRESSPPATSSSPSGGGQGT--LYGCGGAGMTGFTTCQ 225

QY 284 SGTGCKLNDYISQCL 299  
| : : : : :  
| : : : : :  
Db 226 ADGTCKVQNMWYSQCL 241

## RESULT 5

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I precursor - fungus (Trichoderma reesei)  
N:Alternate names: exo-cellulohydrolase I  
C:Species: Trichoderma reesei  
C:Date: 03-Aug-1984 #sequence\_revision 03-Aug-1984 #text\_change 16-Jul-1999

C:Accession: A00902  
R:Shewmaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, B./Technology 1, 691-696, 1983  
A:Title: Molecular cloning of exo-cellulohydrolase I derived from Trichoderma reesei  
A:Reference number: A00902  
A:Accession: A00902  
A:Molecule type: DNA  
A:Residues: 1-513 <SHO>  
A:Experimental source: strain 127  
C:Comment: This is the most abundantly produced cellulase in this filamentous fungus; in

C:Genetics:  
A:Gene: CBH1  
A:Introns: 154/2; 386/3  
C:Function:  
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to  
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain from  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-513/Product: cellulose 1,4-beta-cellulobiosidase I #status predicted <MAT>  
F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.2%; Score 219.5; DB 1; Length 513;  
Best Local Similarity 26.9%; Pred. No. 2.7e-08;  
Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

QY 16 ELVSAASGSGSTRYWDCCKPSGAPGAASQPV-YACDANFRLSDFNWQSGCNG 73  
| : : : : :  
| : : : : :  
Db 211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALTPECTTVGGEICE---GDGC--G 261  
74 SAY-----C---ADQPMVNDNLAYGPAATSIAGSSSSGCCACVALLFTSGPVA 122  
262 GTISDNRYGCTCDPGCCMNPYRLGNTSFYG-----PGSS-----FTLDTK----- 303  
QY 123 GKTWVVGSTSTGDLG---SNQFDIAMPGGVGIENG-----CS---SDFGLPQAQ 168  
| : : : : :  
| : : : : :  
Db 304 -KLTVTQFETSGAINRYVQNGVTFEQPMALISYSGNELNDYCTAEAEAEFGSSPSD 362  
169 YGSISSRDQDSFPAPLKPQCGMRFMFQGN---ADNPTFQQVQCPAEIYASGCKRND 225  
| : : : : :  
| : : : : :  
Db 363 KGGILQFKKATSGGAVLWMSL-WD-DYANMLMLDSTYPTNERTSTPGAV--RSGCSTSS 418  
226 -----DSSEF-----VTFP-----PSGDN---GGTGTPTNAPGSGQTSRGGSS 261  
419 GVPAYVESGSPAKATPNSNIFGISTGNPBGNGPRGNGTTTRRPA---TTGSSP 475  
QY 262 GCTSQKMAQCGGIGSGGCTTVCAGTTCQKLNDRYSQCL 299  
| : : : : :  
| : : : : :  
Db 476 GPTQSHYQCGGIGSGGPTVCASGTTQVILNPIYSQCL 513

## RESULT 6

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma koningi)  
C:Species: Trichoderma koningi  
C:Date: 20-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999

C:Accession: S45380  
R:Wey, T.T.; Hseu, T.B.; Huang, L.  
Curr. Microbiol. 28, 31-39, 1994  
A:Title: Molecular cloning and sequence analysis of the cellobiohydrolase I gene from Tr  
A:Reference number: S45380; MUID:94100788; PMID:7764306  
A:Accession: S45380

A:Molecule type: DNA  
A:Residues: 1-513 <MBY>  
A:Cross-References: EMBL:X69976; NID:9457422; PIDD:CAA49596.1; PID:9457423  
C:Genetics:  
A:Introns: 154/2; 386/3  
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.2%; Score 219.5; DB 2; Length 513;  
Best Local Similarity 26.9%; Pred. No. 2.7e-08;  
Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

QY 16 ELVSAASGSGSTRYWDCCKPSGAPGAASQPV-YACDANFRLSDFNWQSGCNG 73  
| : : : : :  
| : : : : :  
Db 211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALTPECTTVGGEICE---GDGC--G 261  
74 SAY-----C---ADQPMVNDNLAYGPAATSIAGSSSSGCCACVALLFTSGPVA 122  
262 GTISDNRYGCTCDPGCCMNPYRLGNTSFYG-----PGSS-----FTLDTK----- 303  
QY 123 GKTWVVGSTSTGDLG---SNQFDIAMPGGVGIENG-----CS---SDFGLPQAQ 168  
| : : : : :  
| : : : : :  
Db 304 -KLTVTQFETSGAINRYVQNGVTFEQPMALISYSGNELNDYCTAEAEAEFGSSPSD 362  
169 YGSISSRDQDSFPAPLKPQCGMRFMFQGN---ADNPTFQQVQCPAEIYASGCKRND 225  
| : : : : :  
| : : : : :  
Db 363 KGGILQFKKATSGGAVLWMSL-WD-DYANMLMLDSTYPTNERTSTPGAV--RSGCSTSS 418  
226 -----DSSEF-----VTFP-----PSGDN---GGTGTPTNAPGSGQTSRGGSS 261  
419 GVPAYVESGSPAKATPNSNIFGISTGNPBGNGPRGNGTTTRRPA---TTGSSP 475  
QY 262 GCTSQKMAQCGGIGSGGCTTVCAGTTCQKLNDRYSQCL 299  
| : : : : :  
| : : : : :  
Db 476 GPTQSHYQCGGIGSGGPTVCASGTTQVILNPIYSQCL 513

## RESULT 7

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma viride)  
C:Species: Trichoderma viride  
C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1999

C:Accession: S11439  
R:Cheng, C.; Tsukagoshi, N.; Ueda, S.  
Nucleic Acids Res. 18, 5559, 1990  
A:Title: Nucleotide sequence of the cellobiohydrolase gene from Trichoderma viride.  
A:Reference number: S11439; MUID:91016856; PMID:2216737  
A:Accession: S11439  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-513 <CEB>  
A:Cross-References: EMBL:X53931; NID:95196; PIDD:CAA37878.1; PID:9299397

C:Genetics:  
A:Introns: 154/2; 386/3  
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 13.0%; Score 215.5; DB 2; Length 513;  
Best Local Similarity 26.2%; Pred. No. 5.2e-08;  
Matches 86; Conservative 36; Mismatches 137; Indels 69; Gaps 16;

QY 16 ELVSAASGSGSTRYWDCCKPSGAPGAASQPV-YACDANFRLSDFNWQSGCNG 73  
| : : : : :  
| : : : : :  
Db 211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALTPECTTVGGEICEGDSGGTTS 266  
74 SAY---SC---ADQPMVNDNLAYGPAATSIAGSSSSGCCACVALLFTSGPVAKTMV 127  
267 DRYGCTCDPGCCMNPYRLGNTSFYG-----PGSS-----FTLDTK-----KLT 307  
QY 129 VQSTSTGDLG---SNQFDIAMPGGVGIENGCS-----SDFGLPQAQYGGIS 173  
| : : : : :  
| : : : : :

Db 308 VTQETSAINRYVQNGVTEQQPNAELGDSGNSLDDYCAEABFRCSSSFDKGGTL 367  
Qy 174 SRDCCSPAPLKECCQMRPDPQN---ADNPTFPQVCCPAEI---VARSCCKAND 226  
Db 368 QFKKATSGAMVWML-WD-DYYANMLMDSTYPTDETSTPGAVGSSSTSSVPAQLE 425  
Qy 227 SSFP-----VFTP-----PSGNGSTGTPSTAPSGGTSGGSGSCTGQKAC 271  
Db 426 SNSPNKAVVYSNIRKPIGSGNGGPNPPTTPPATSTGSSPGTGTTHGQC 485  
Qy 272 GGIGSGCTTGVSGTTCOKLNDYSSQL 299  
Db 486 GGIGTPTVCASSSTCCYLAFFYSQCL 513

## RESULT 8

cellulase (EC 3.2.1.4) A precursor - fungus (Trichoderma reesei)  
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I  
C:Species: Trichoderma reesei  
C:Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 22-Jun-1995  
C:Accession: A25928; A25565  
R:Van Arsdell, J.N.; Kwock, S.; Schweickart, V.L.; Ladner, K.B.; Gelfand, D.H.; Innis, M.  
Bio/Technology 5, 60-64, 1987  
A:Title: Cloning, characterization, and expression in Saccharomyces cerevisiae of endogly-  
A:Reference number: A25928  
A:Accession: A25928  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-459 <VAN>  
A:Experimental source: strain L27  
R:Penitella, M.; Lehtovaara, P.; Nevalainen, H.; Birkhahn, R.; Knowles, J.  
Gene 45, 253-263, 1986  
A:Title: Homo-cy between cellulase genes of Trichoderma reesei: complete nucleotide seq-  
A:Reference number: A25565; MUID:87106822; PMID:2948877  
A:Accession: A25565  
A:Molecule type: mRNA  
A:Residues: 1-459 <PEN>  
A:Cross-references: GB:M15665; NID:GL70546; PIDD:AAA34212.1; PID:GL70547  
C:Genetics:  
A:Gene: eg11  
A:Introns: 257/2; 454/1  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce-  
A:Pathway: cellulose degradation  
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain hom-  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-459/Product: cellulase A #status predicted <MAT>  
F:428-459/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.9%; Score 214; DB 2; Length 459;  
Best Local Similarity 27.0%; Pred. No. 6e-08;  
Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

Qy 19 ASAASGSGS---TRWDCRCPKAPKAAVAPYACD-----NQRLSPFN 66  
Db 101 ASGTTSSSLTMNQ-----MPSSS-OGYSVSPRLYLDDSDGYMLKLNQELS FDY 154  
Qy 67 Q-----SGNGSAYSCADQTP--NAVNDNIANGFA 95  
Db 155 DLSALPGENGSLYLSQMDENGANGYNTAGANGSGY-CAQCPVQWTR-NGTL----- 207  
Qy 96 KTSIAGSSESSMCC-----AC-----YALFTS--GPV 121  
Db 208 -----NTSHQFCPCNEMDILBGNRANALTPHSCTATACDSAGCGFNPYSGYKSYGP- 261  
Qy 122 AGKTVVQST-----STGDLGS-----NQPD1--AMPGGVGIFFNGCSQ 163  
Db 262 -GTVDSKFTFTITPNDNGSPSGNLVITRKQCGVCIPEAGGGDT--ISSCS- 317  
Qy 161 FGGI-PGAQYGGISSRQCDSPAPLPGCQWRT-WFQND--NPTFTQVQC----- 211

Db 318 -----ASAYGLATWGA-----LSSGAVLVFSIMWNSQYMMWLDGNAFGCSSTEGN 366  
Qy 212 PSEIYASGCKRRDSDFFVTPPSGNGGTPSTTA-----PSGGT-----SPGG 259  
Db 367 PSNIIA-----NNPHTVVFSSNIRWMDIS--TTNSTAPPPPPASSTFTTTRSSITSS 419  
Qy 260 GSGCTGKMAQCGGIGFSCTTGVSGTTCOKLNDYSSQL 299  
Db 420 SPSCQTQTHMGCGGIGSGCKTCTSGTTQYXNDYSSQL 459

## RESULT 9

cellulase (EC 3.2.1.4) - fungus (Trichoderma longibrachiatum)  
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I homolog EGL  
C:Species: Trichoderma longibrachiatum  
C:Date: 19-Nov-1993 #sequence\_revision 18-Nov-1994 #text\_change 22-Jun-1995  
C:Accession: A48375; S28521  
R:Gonzalez, R.; Ramon, D.; Perez-Gonzalez, J.A.  
Appl. Microbiol. Biotechnol. 38, 370-375, 1992  
A:Title: Cloning, sequence analysis and yeast expression of the eg11 gene from Trich-  
A:Reference number: A48375; MUID:93159747; PMID:1369161  
A:Accession: A48375  
A:Contents: CECT 2606  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-463 <GON>  
A:Cross-references: EMBL:X60652; NID:95181; PIDD:CAA43059.1; PID:95182  
A>Note: sequence extracted from NCBI backbone (NCBIN:125157, NCBI:P.125158)  
C:Genetics:  
A:Introns: 257/2; 458/1  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as  
A:Pathway: cellulose degradation  
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h  
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation  
F:432-463/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.2%; Score 203; DB 2; Length 463;  
Best Local Similarity 28.4%; Pred. No. 3.6e-07;  
Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

Qy 68 SGNGSAYSCADQTP--NAVNDNL---AVGPAAT--SIAGSSSSWCCACTALPTSG 119  
Db 184 AGANGSGY-CAQCPVQWTR-NGTLNTPSGGFCNEMDILBGNRANALTPHSCTATAC 241  
Qy 120 PVAG-----KTMV-----QSTGDLGS-----NQPD 143  
Db 242 DSAAGGFNPYSGYPNYFGGDTVTSSTFTITQFNTPNDSPSGNLVITRKTRQNGVD 301  
Qy 144 I--AMPGGVGIFFNGCSQFGGLPGAQYGGISSRDQDSFPAPLPGCQWRT-WFQND 200  
Db 302 IPEAKGGDT--ISSCS-----ASAYGLATWGA-----LSSGAVLVFSIMWNSQ 347  
Qy 201 -----NPTFTQVCCPAIYASGCKRRDSDFFVTPPSGCGN----- 239  
Db 348 YMMWLDGNAFGCSSTEGNPSNII--ANNPHTVVFSSNIRWMDIS--TTNSTAGAPP 402  
Qy 240 ---GCTGPTATAGSGQSPGSSGCTGKMAQCGGIGFSCTTGVSGTTCOKLNDYSS 296  
Db 403 PEPASSTFTTTRSSITSS--SSPSCQTQTHMGCGGIGTGCCTCTSGTTQYXNDYSS 460  
Qy 297 QCL 299  
Db 461 QCL 463

## RESULT 10

S33164  
cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chrysos-  
C:Species: Phanerochaete chrysosporium  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Jun-1995  
C:Accession: S33164

Best local similarity 2071; gaps not used for  
Matches 102; Conservative 32; Mismatches 108; Indels 160; Gaps 25

QY 266 QKNAQCGGIGFSGCTTCVSGTTCQKANEYISQCL 299  
| | | | : : | | | | | : : | | | |  
DB 269 THNGQCGGQGWGPTQCESGTTCCVSIQWMSQCL 302

## RESULT 13

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (Humicola grisea)

N/Alternate names: beta-glucanocellobiohydrolase; exoglucanase

C/Species: Humicola grisea var. thermoides

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: S38794; S08240; A45869

R/Radford, A.

submitted to the EMBL Data Library, June 1991

A/Reference number: S38794

A/Accession: S38794

A/Molecule type: DNA

A/Residues: 1-525 &lt;RAD&gt;

A/Cross-References: EMBL:X17258; NID:g2760; PDB:CAA3150.1; PID:g2761

A/Note: this is a revision to the sequence from reference S08240

R/de Oliveira Azevedo, M.; Radford, A.

Nucleic Acids Res. 18, 668, 1990

A/Title: Sequence of cbb-1 gene of Humicola grisea var. thermoides.

A/Reference number: S08240; MUID:90175006; PMID:2308855

A/Accession: S08240

A/Molecule type: DNA

A/Residues: 1-299, 'H', 301-525 &lt;DEO&gt;

A/Cross-References: EMBL:X17258

A/Note: The authors translated the codon CAG for residue 87 as His

R/Azevedo, M.; de O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.

J. Gen. Microbiol. 136, 2569-2576, 1990

A/Title: Cloning, sequencing and homologies of the cbb-1 (exoglucanase) gene of Humicola

A/Reference number: A45869; MUID:91178527; PMID:2127803

A/Accession: A45869

A/Status: not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQR', 241-244, 'T

A/Cross-References: GB:M64588; GB:X17258

A/Note: this sequence has been revised. See entry S08240

C/Genetics:

A/Genes: cbb-1

A/Insertions: 138/2

C/Supernames: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain hom

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

F/494-525/Domain: fungal cellulose-binding domain homology &lt;FCB&gt;

Query Match 11.2%; Score 185; DB 1; Length 525;

Best Local Similarity 24.9%; Pred. No. 7.5e-06;

Matches 97; Conservative 33; Mismatches 136; Indels 124; Gaps 23;

Db 11 LAALPLVSAASGSGSTRYWDCCKPSCAMPK-KAAVSQPVYACDANFORLSDP----- 64

Db 159 LAGLALYVSMADGG--LSRY-----PKNAGAKGTGYCDACQPRIDKXINGEA 206

Db 65 NVSGGCGN-----GGSAYS--CADOTPAVNDLAVGEA--ATSIAGGS--ESSKCCAC 111

Db 207 NIEGWTGSTNDPNAAGRYGTCCSEMDIWEAN-NMAATFPHPCITIIIGSRCEGDSGCGT 265

Db 112 YALFTTSG-----PVAGKTWVYOST-----STGGDLGS--NOR 142

Db 266 YSNERYAGVDDPPGCDPNSYKQKTFYKGTGTTVTYTKITVTYGTCLKAKNCDLGEIKER 325

Db 143 -----DIAMPG--GQVGFNGCASSQ--FGGLPG-AQYGGISRRQCDSPAPF 185

Db 326 YVQDGKTIIPNSSTIPGVEKSTITQWCDQKVAAGDIDFNKKGKQMKALAGPMVL 385

Db 186 KPGQGRFQFQANPNF---TFQVQCPATIVASCCRN---DDSEFPV----- 231

Db 366 VMSI-----NDHSAANMLDSTF-----PVDAKGPFAERGACTTSVPEVEADAPN 435

Db 232 -----FTP-----PSGGNGGTG-----TPTSTAPSGSGTSPGSGGCTSOANA 269

Db 436 SNVFSNIRGPIGTAGVAGAGAGGNNKGNKAPPTTTTSSAPATTTAAGTKAGRW 495

Db 270 QCGGIGFSGCTTCTVSGTTCCQKNDYISQCL 299

Db 496 QCGGIGFTGPTQCEEPYICTKNDYISQCL 525

## RESULT 14

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I - basidiomycete (Phanerochaete chrysosporium)

C/Species: Phanerochaete chrysosporium

C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Sep-1998

C/Accession: J50083

R/Sims, P.; James, C.; Broda, P.

Gene 74, 411-422, 1988

A/Title: The identification, molecular cloning and characterisation of a gene from Ph

A/Reference number: J50083; MUID:89232732; PMID:3246351

A/Accession: J50083

A/Molecule type: DNA

A/Residues: 1-516 &lt;SIM&gt;

A/Experimental source: strain ME446

C/Comment: This protein is a component of the inducible cellulase complex.

C/Genetics:

A/Insertions: 201/3; 511/1

C/Supernames: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

F/485-516/Domain: fungal cellulose-binding domain homology &lt;FCB&gt;

F/81,92/Active site: Asp #status predicted

Query Match 11.1%; Score 184; DB 2; Length 516;

Best Local Similarity 24.7%; Pred. No. 8.7e-06;

Matches 86; Conservative 37; Mismatches 113; Indels 112; Gaps 20;

Db 20 SAASGSGSTRYWDCKPSCAMPK-KAAVSQPVYACDANFORLSDPVSQNGG----- 73

Db 211 SANAGTG---NYGCTCEMDIWEANDMAAYTE-HPCIT-----NAQTRCSGSDCTR 258

Db 74 -----SAYSCADOTPAVNDLAVGEA--ATSIAGGS--ESSKCCACVATFTSPGV 121

Db 259 DTGLCADAGCDPNSYKQKTFYKGTGTTVTYTKITVTYGTCLKAKNCDLGEIKER 306

Db 122 AG-----KTVWVSGTSTGDLGSLQPDAMPG-GGVGIF--NGCSAQ--FGGLP-GAQY 169

Db 307 AGTLLEIRLRYQN-----QKVIQNS-SVKIPGIDPNSITDNFSGQKXAFGDTNFAQH 361

Db 170 GGISRRQCGN-----DSPAPDKPCQMRDWFQ-----NADNTFTFQVQC- 211

Db 362 GGLKQVGEALRTGQVLAISIMDYAAMNL-----WLSNYFTNDPSTPGVAGGTGA 413

Db 212 -----PARIYASGCGRNDSSFPVTPSPGSGNGT-----GTPSTAP 250

Db 414 TTSVPADIERQS-----PNAIVFSNIRGPIGTAGVAGAGGNNKGNKAPPTTTTSSAP 467

Db 251 GSGQTSPPGSGGCTSOANACCGGIGFSGCTTCTVSGTTCCQKNDYISQCL 298

Db 468 SSSSTPPTQPTGVTVPWQCGGIGYGTTCASPYTCVAVNPYSQCL 515

## RESULT 15

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I - Penicillium janthinellum

N/Alternate names: exo-cellobiohydrolase

C/Species: Penicillium janthinellum

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C/Accession: J00150; S21508

R/Christoph, A.X.; Weigel, T.O.; Schulte, G.

Gene 124, 57-65, 1993

A/Title: Cloning, sequencing, and heterologous expression of a cellulase-encoding cDN

A/Reference number: J00150; MUID:93178976; PMID:8440481

A/Accession: J00150

A/Molecule type: mRNA

A/Residues: 1-537 &lt;CHR&gt;

A/Cross-References: EMBL:X59054; NID:g3177; PDB:CAA41780.1; PID:g3178

A/Note: the authors translated the codon ATG for residue 172 as Asn

C/Genetics:

A/Genes: cbb1

C/Supernames: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:506-537/Domain: fungal cellulose-binding domain homology <FCB>

Query Match: 11.04; Score 183; DB 1; Length 537;

Best local similarity 23.44; Pred. No. 1.1e-05;

Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;

```
QY 24 GSQSLRITWDCCKRSCAMPKAAVSQPIYACDANFQRLSDRYVSGCNGSAYSACADQTP 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 GYSLIDRYGGTCDEP-----GCDFFPYRMGVTFY-----GGRTIDTKSP 308

QY 84 MAY-----NDNLAYGPAA-----TSIAGGS-ESSWCCACVALTF 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 FVIVVQFGLNDGISTGLSEIKRFYVGGKVTIGNPQSTIVGSGNSITDSVCNA----- 362

QY 117 TSGPVAKMTWVGSTGTGDLGNOFDI--AMPGGVGIENGQ----- 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 -----QKSAFGD--TNEFSKGGMAGMAGLADGMVIVMSLWDDHASDMLWL 407

QY 158 -----SSQPGGLPGACYGGLSSRDQDSFPAFLKPGQWRFPQXADNPFTTFQOVQCP 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 DSTYPTNATSTTPGAK-----RGTCDSLRRP-----NTVESTY-----P 441

QY 213 AEIVARSGCKRNDSSPVPFPPSGANGGTGTPSTAPGSGQ-----TSPGGGS 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 NAYVIYSNKTGPIINS--FTGGTSSSSSTTTTSKSTSTSSSKTTTIVTTTSSSSS 499

QY 262 GCTSQMMAQCGGIGFGGCTTCVSGITTCQKINDYISQCL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 500 GTGARDMAQCGGNGMTGPTCVSPYTCTKQNDWYSQCL 537
```

Search completed: July 7, 2004, 08:54:54  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 08:46:50 ; Search time 17 Seconds

(without alignments)  
915,822 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSTPVLRKTLAAAPLVAS.....TCVSGTTCQKNDYSGQL 295

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	942.5	56.8	1 GUNK_FUSOX	P45699 fusarium ox
2	921.5	55.5	1 GUN5_HUMIN	P43316 humicola in
3	507	30.6	1 GUNB_PSEFL	P18126 pseudomonas
4	462	27.8	1 GUN1_USAMA	P54424 usiliago ma
5	247	14.3	1 PABP_PENNY	Q98e18 penicillium
6	221	13.3	1 GUN5_TRIPE	P43317 trichoderma
7	219.5	13.2	1 GUN1_TRIPE	P00725 trichoderma
8	215.5	13.0	1 GUN1_TRIPE	P19355 trichoderma
9	203	12.9	1 GUN1_TRIPE	P07981 trichoderma
10	203	12.2	1 GUN1_TRIPE	Q12714 trichoderma
11	185	11.2	1 GUN1_HUMET	P15828 humicola gr
12	184	11.1	1 GUN1_PENAC	P15860 phanerochaete
13	183	11.0	1 GUN1_NEUCR	Q06886 neurospora
14	181	10.9	1 GUN1_AGABI	P00023 agaricus bi
15	172	10.4	1 GUN1_AGABI	Q14405 trichoderma
16	159	9.6	1 GUN4_TRIPE	P46238 fusarium ox
17	158.5	9.6	1 GUN4_TRIPE	P46238 fusarium ox
18	156	9.4	1 GUN2_AGABI	Q92400 agaricus bi
19	149.5	9.0	1 GUN1_ASAPC	Q05984 aspergillus
20	143.5	8.6	1 GUN3_HUMIN	Q12664 humicola in
21	137	8.3	1 GUN3_AGABI	P49075 agaricus bi
22	137	7.8	1 GUN2_TRIPE	P07982 trichoderma
23	127	7.7	1 GUNF_FUSOX	P46239 fusarium ox
24	127	7.7	1 GUNB_FUSOX	P46236 fusarium ox
25	127	7.7	1 GUNB_FUSOX	Q9v6b9 drosophila
26	125.5	7.6	1 GUN4_DROME	Q9v6b9 drosophila
27	124	7.5	1 GUN2_TRIPE	P21490 homo sapien
28	122	7.3	1 KRUC_SHEEP	P26312 ovine atles
29	120.5	7.2	1 34KD_MYCTU	P15566 mycobacteri
30	120	7.2	1 ZAN_MOUSE	Q88729 mus musculu
31	118.5	7.1	1 PSBP_PORPU	P50272 porphyria pu
32	118.5	7.1	1 VG_DROME	Q26366 drosophila
33	118.5	7.1	1 TROP_HUMAN	Q12816 homo sapien

34	113.5	6.8	2090	1	N214_HUMAN	P35658 homo sapien
35	110.5	6.7	675	1	YMW2_GAEEL	P34504 canorhabdi
36	109.5	6.6	518	1	RUS_MOUSE	P56859 mus musculu
37	109	6.6	660	1	YHLI_BBY	P03181 epstein-bar
38	107	6.4	967	1	ATSI_PAT	Q9wq1 ratius norv
39	107	6.4	968	1	ATSI_MOUSE	P97857 mus musculu
40	106.5	6.4	2704	1	G168_PAPR	P17053 parametium
41	105.5	6.4	452	1	GUX1_CRYPA	Q00548 cryphonectr
42	105	6.3	481	1	LORI_MOUSE	P18165 mus musculu
43	103	6.2	394	1	THID_CLAFS	Q9u14 claviceps f
44	103	6.2	491	1	YK98_MYCTU	Q10707 mycobacteri
45	103	6.2	537	1	SP70_DICDI	P15269 dictyostell

## ALIGNMENTS

RESULT 1	ID	Accession	Standard	Protein	Length
AC	P45699	GUNK_FUSOX	STANDARD	PRT	376 AA.
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).				
OS	Fusarium oxysporum.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Mitosporic Hypocreales; Fusarium.				
OX	NCBI_Taxid=5507;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95047531; PubMed=7959045;				
RA	Shepard P.O., Grant F.J., Oort P.T., Sprecher C.A., Foster D.C.,				
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.D.;				
RT	"The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from Fusarium oxysporum.";				
RT	Gene 150:163-167 [1994].				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.				
CC	-1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.				
CC	-1- SIMILARITY: Belongs to cellulase family K (family 45 of glycoyl hydrolases).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL, L29381; AAA65589.1; -.				
DR	HSSP, P43316; ZENG.				
DR	InterPro, IPR009009; Barwin like.				
DR	InterPro, IPR000254; CBD fungal.				
DR	InterPro, IPR000334; Glyco_hydro_45.				
DR	Pfam, PF00734; CBM_1; 1.				
DR	Pfam, PF02015; Glyco_hydro_45; 1.				
DR	SMART, SM00236; FCBP; 1.				
DR	PROSITE, PS00562; CBD_FUNGAL; 1.				
DR	PROSITE, PS01140; GLYCOSYL_HYDROL_P45; 1.				
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
KW	SIGNAL				
FT	CHAIN	19	376		POTENTIAL.
FT	CHAIN	19	308		POTENTIAL ENDOGLUCANASE TYPE K.
FT	DOMAIN	309	338		CATALYTIC.
FT	DOMAIN	339	376		LINKER.
FT	DOMAIN	29	29		CELLULOSE-BINDING.
FT	ACT_SITE	29	29		NUCLEOPHILE (BY SIMILARITY).
FT	ACT_SITE	140	140		PROTON DONOR (BY SIMILARITY).
FT	SEQUENCE	376 AA;	39235 MW;		P430A5F62B9F882 CMC64;
Query Match	56.8%	Score 942.5;	DB 1;	Length 376;	

Best Local Similarity 49.5%; Pred. No. 1e-61;  
Matches 183; Conservative 36; Mismatches 72; Indels 79; Gaps 6;

QY 9 TTTAALPLVWASAGSGGSTRYWDCCPKSCAMPKAAVSQPVYACDANFQRLSDFNVD 68  
DB 5 TLTLALAPLWASAGSGHSTRYWDCCPKSCAGKAAVAPALTCDDKNDPISNTNANV 64  
QY 69 GC-NGGSAVSCADQTPANVNDINAFPAATSIAGGSSSSGCCACALPFTSGPAGKTV 127  
DB 65 GCGGGSAVACTNTPMAVNDLAVGFPAATISGSBASWCACALPFTSGPAGKTV 124  
QY 128 VQSTGTGDLGSDNEDLAMPGGVGI-FNGCSSQFG-GLPAPAGVIGISSRDQSDFPAPK 186  
DB 125 VQSTGTGDLGSDNEDLAMPGGVGI-FDQCTSFPAKALGAGVIGISSRSCDYPALIK 184  
QY 187 PGQWRFPWFQNAADNPFTTQVQVCEALIVASGCKNDSSFPVF----- 232  
DB 185 DQGWRFDFENADNPDETEQVQCEKALDISCKDDSSFPAFKADTSKSPSS 244  
QY 233 -----TPPS-----GNGCTGTPTSPAGSGQTS 257  
DB 245 AKKTTAAAAQPQTKDAPVQKSTPAAQEPKPADKFPQTDKPAKPAKTPAQ 304  
QY 258 GGGSGCTSOX-----MAQCGIGF--SGCTTCVSGTTCQ 289  
DB 305 PVKPKTTQKVGKTRGSCPAKTDATAKASVPAVYQCGSKAVPQGNLACATGSKCV 364  
QY 290 KLDYISQCL 299  
DB 365 KQNEYYSQCV 374

RESULT 2  
GUNS\_HOMIN  
ID GUNS\_HOMIN STANDARD; PRT; 213 AA.  
AC P43316;  
DT 01-NOV-1995 (Ref. 32, Created)  
DT 01-NOV-1995 (Ref. 32, Last sequence update)  
DT 10-OCT-2003 (Ref. 42, Last annotation update)  
DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)  
DE (Cellulase V) (EG V).  
OS Humicola insolens.  
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
NCBI\_TaxID=34413;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rasmussen G., Mikkelson J.-M., Schuelein M., Packar S.A., Hagen F.,  
RT Hjord C.M., Hastrup S.;  
RL "A cellulase preparation comprising an endoglucanase enzyme";  
RL Patent number WO9117243, 14-NOV-1991.  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RA MEDLINE=93390621; PubMed=8377830;  
RA Davies G.-C., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,  
RT Wilson K.S., Hjord C., Mikkelson J.M., Rasmussen G., Schuelein M.;  
RL "Structure and function of endoglucanase V";  
RL Nature 365:362-364 (1993).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RA MEDLINE=96101453; PubMed=8519779;  
RA Davies G.-C., Tolley S.P., Hentrisat B., Hjord C., Schuelein M.;  
RT "Structures of oligosaccharide-bound forms of the endoglucanase V  
from Humicola insolens at 1.9-Å resolution.";  
RL Biochemistry 34:16210-16220 (1995).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
RA Davies G.-C., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,  
RT Wilson K.S., Rasmussen G., Schuelein M.;  
RL "Structure determination and refinement of the Humicola insolens  
endoglucanase V at 1.5-Å resolution.";  
RL Acta Crystallogr. D 52:7-17 (1996).  
CC -/- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.

CC -/- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl  
hydrolases).  
DR PDB; 2ENG; 08-DEC-96.  
DR PDB; 3ENG; 16-JUN-97.  
DR PDB; 4ENG; 16-JUN-97.  
DR PDB; 1HD5; 03-JUL-01.  
DR InterPro: IPR009009; Barvin like.  
DR InterPro: IPR000334; Glyco hydro 45.  
DR Pfam: PF02015; Glyco hydro 45; 1.  
DR PROSITE; PS01140; GLYCOSTYL\_HYDROL\_F45; 1.  
KW PROSITE; Glycolase degradation; Hydrolase; Glycosidase; 3D-structure.  
FT ACT\_SITE 10 10  
FT ACT\_SITE 121 121  
FT STRAND 2 8  
FT STRAND 12 12  
FT HELIX 15 17  
FT TURN 19 20  
FT STRAND 24 24  
FT STRAND 31 31  
FT TURN 33 34  
FT STRAND 37 37  
FT TURN 41 42  
FT STRAND 45 45  
FT TURN 46 47  
FT TURN 49 50  
FT STRAND 53 54  
FT TURN 57 58  
FT STRAND 62 65  
FT TURN 66 67  
FT STRAND 68 76  
FT TURN 78 79  
FT HELIX 82 85  
FT TURN 86 87  
FT STRAND 89 94  
FT HELIX 97 99  
FT TURN 100 101  
FT STRAND 103 110  
FT STRAND 119 123  
FT TURN 125 126  
FT TURN 130 131  
FT HELIX 135 139  
FT STRAND 145 145  
FT TURN 146 148  
FT STRAND 149 149  
FT HELIX 153 158  
FT HELIX 161 163  
FT HELIX 164 171  
FT TURN 172 173  
FT TURN 175 176  
FT STRAND 181 187  
FT HELIX 191 197  
FT STRAND 201 201  
FT TURN 202 203  
FT HELIX 204 206  
SQ SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;

Query Match 55.5%; Score 921.5; DB 1; Length 213;  
Best Local Similarity 74.4%; Pred. No. 1.9e-60;  
Matches 157; Conservative 27; Mismatches 26; Indels 1; Gaps 1;

QY 26 GQSTRYWDCCPKSCAMPKAAVSQPVYACDANFQRLSDFNVDGCG-NGGSAVSCADQTPW 84  
DB 3 GSTRYWDCCPKSCAMPKAAVSQPVYACDANFQRLSDFNVDGCG-NGGSAVSCADQTPW 62  
QY 85 AVNDINLAFPAATSIAGGSSSSGCCACALPFTSGPAGKTVVQSTGTGDLGSDNED 144  
DB 63 AVNDINLAFPAATSIAGGSSSSGCCACALPFTSGPAGKTVVQSTGTGDLGSDNED 122  
QY 145 AMPGGVGI-FNGCSSQFG-GLPAPAGVIGISSRDQSDFPAPKLPKGGCOMFQNAADNPFT 204  
DB 123 NITGGVGI-FDQCTSFPAKALGAGVIGISSRSCDYPALIKPGCVWRFQNAADNPFT 182  
QY 205 TFOQVQCEALIVASGCKNDSSFPVFPTTP 235



FT CHAIN 27 393 ENDOLUCINASE 1.  
 FT ACT\_SITE 34 34 NUCLEOPHILE (BY SIMILARITY).  
 FT ACT\_SITE 152 152 PROTON DONOR (BY SIMILARITY).  
 FT DOMAIN 270 385 ALA/GLY/SER-RICH.  
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 393 AA; 39594 MW; 65C753C610C6AD3 CRC64;  
 Query Match 27.8%; Score 462; DB 1; Length 393;  
 Best local similarity 30.9%; Pred. No. 9,3e-27;  
 Matches 115; Conservative 41; Mismatches 100; Indels 116; Gaps 12;  
 QY 11 LAALPLVMSAASGSGSTGYMDCKSCAPGKAASQPVYACANQRL---SDPNV 66  
 DB 12 LSLSLVHLGDEVRAGMATRYMDCCASISWEGAPVAVDNCADVTLLSKDPG 71  
 QY 67 QSGNGSAYSCADQTPW--AVNDNLAYGFATSIAGSSSSWCCALYFTTSGPVAK 124  
 DB 72 QSGNGGNGKFKSCMGPFDEMDPTLAFRFGA--FTTGQSDPDGCAVYAF--BHAQK 128  
 QY 125 TM---VVOSTSTGDLGNSQFDLAMPGGVGF--NGSSQFG--GLPGAQYGISR 175  
 DB 129 AMKNTLIFQVTVNGGVQSQMFQIPGSGIGAFPKGCPAGWYEAALMGDQYGVKSA 188  
 QY 176 DQDSPFAPLXPGQOMRF--WFMADNPTF--TQVQCGAFIYARSGCKNDSSPVF 232  
 DB 189 TEGSKLPKPLQSGCKMRSE--GDNPLKSGPRKVCCKSLIDRSQGRKNDITSPY 245  
 QY 223 T----- 233  
 DB 246 SKAYDSANTAPAYKDRGVCLAGGKSKSAAGVYDGSADAGADAGAGAA390G 305  
 QY 234 -----PESGN-----GCTPTPTSPSGGTSFGG 260  
 DB 306 QPBGYQPPSGNDQSGSWGATGTAGSSGSDSGSTANGSGSGAPSDSAAVAPPSG 365  
 QY 261 SGGTSQKMAQCG 272  
 DB 366 SNPPAAQGGQGG 377  
 RESULT 5  
 FAEB PENFN STANDARD; PRT; 353 AA.  
 AC Q9HRI8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Feruloyl esterase B precursor (BC 3.1.1.73) (Ferulic acid esterase B)  
 DE (FAEB) (Citramoyl esterase).  
 GN FAEB.  
 OS Penicillium funiculosum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
 OX NCBI\_TaxID=28572;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 19-68; 90-97; 163-167; 212-241 AND  
 RP 253-274, FUNCTION, SUBCELLULAR LOCATION, AND INDUCTION.  
 RC STRAIN=IMT 134756;  
 RX MEDLINE=20534747; PubMed=11082184;  
 RA Kroon P.A., Williamson G., Fish N.M., Archer D.B., Belshaw N.J.;  
 RT "A modular esterase from penicillium funiculosum which releases  
 RT ferulic acid from plant cell walls and binds crystalline cellulose  
 RT contains a carbohydrate binding module.";  
 RL Eur. J. Biochem. 267:6740-6752(2000).  
 CC -!- FUNCTION: involved in degradation of plant cell walls.  
 CC Hydrolyzes the feruloyl-arabinose ester bond in arabinoxylans, and  
 CC the feruloyl-galactose and feruloyl-arabinose ester bonds in  
 CC pectin. Binds strongly to cellulose.  
 CC -!- CATALYTIC ACTIVITY: Feruloyl-polysaccharide + H(2)O = ferulate +  
 CC polysaccharide.  
 CC -!- ENZYME REGULATION: Inhibited by the specific serine esterase  
 CC inhibitor AEBSP.  
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- INDUCTION: Repressed by glucose, probably via the carbon  
 CC catabolite repressor protein Crea.  
 CC -!- PTM: Glycosylated (Probable).  
 CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AJ291436; CAC1414.1; -.  
 DR HSBP; P00725; ZCBH.  
 DR GO; GO:0005576; C:extracellular; IDA.  
 DR GO; GO:0030248; F:cellulose binding; IDA.  
 DR GO; GO:0030600; F:feruloyl esterase activity; IDA.  
 DR GO; GO:0016988; P:cell wall catabolism; IDA.  
 DR GO; GO:0045490; P:pectin catabolism; IDA.  
 DR GO; GO:0045493; P:xyilan catabolism; IDA.  
 DR InterPro; IPR000254; CBD\_fungal.  
 DR InterPro; IPR00379; Ser\_estrs.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR Prodom; pnc01823; CBD\_fungal; 1.  
 DR SMART; SM00236; fCBD; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 DR HydroLase; Serine esterase; Xylan degradation; Glycoprotein; Signal.  
 FT CHAIN 1 18  
 FT SIGNAL 1 18  
 FT DOMAIN 19 353 FERULOYL ESTERASE B.  
 FT DOMAIN 292 318 CATALYTIC (PROBABLE).  
 FT DOMAIN 315 353 PRO/SER/THR-RICH (LINKER).  
 FT ACT\_SITE 136 136 CELLULOSE-BINDING (PROBABLE).  
 FT ACT\_SITE 179 179 CARBOHYD CHARGE RELAY SYSTEM.  
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 353 AA; 37333 MW; 4A08FEDDA9B2D9 CRC64;  
 Query Match 14.9%; Score 247; DB 1; Length 353;  
 Best local similarity 27.5%; Pred. No. 3.6e-11;  
 Matches 106; Conservative 28; Mismatches 116; Indels 136; Gaps 21;  
 QY 6 VIRTTLAALPLVMSAASGSGSTGYMDCKSCAPGKAASQPVYACANQRL---VYACANQ 60  
 DB 12 LSLVTLASLTLQVNNQDNPSLQMT-----IYENKLA-SKPAITVAMHC----- 57  
 QY 61 LSDFNWGGGNGSA-----YSCADQ-----TPAVND-NL--AYGFATSIAGG 102  
 DB 58 -----GSALEYGMVDHSPADQYGYLILYPAIRDNCEADAYSSASLTHNGG 106  
 QY 103 SES-----SWCCACTALFTSGPVAKTMMVOSTSTGDLGNSQFDLAMPGGVGFN 155  
 DB 107 SDSLIVMYVMYVISTYG-----ADSSKVTMTSSSGALMTN-----VLAGAVDYF- 153  
 QY 156 GCSGFGELPGA-QYGG-----ISSRQXSPFAPLXPGQCM-----RF 193  
 DB 154 AAGSFFSSMPACLYGAGADPIMSNVTCGQIQ-HYQQAAYVANGPYGTQYREL 212  
 QY 194 D-WFQNDNP-----TFPQVQCPAEIVARSGCKR--NDSSEF 230  
 DB 213 QMHEGTDADVNTSYADLGEISQWTTMGSLFTNGQNTPL-----SGYTRAVYGDSEKFO 267  
 QY 231 VETPPSG-----NGTGTPTSTARSQGTSPGSGGTSQXMAQCG 273  
 DB 268 AYSAGVGHFVPDVSVMWDFRGITSTTTTPTTPTTSTSPSGCTAAHWAQCG 327  
 QY 274 IGFSGCTCVSGTTCOKLNDVYSQCL 299  
 DB 328 IGYSGCTCASPYTCOKANDIYSQCL 353  
 RESULT 6  
 GNS\_TRAIR

ID GUN5 TRIRE STANDARD; PRT; 242 AA.  
AC P43317;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V) (Cellulase V) (EG V).  
GN EG-5.  
OS Trichoderma reesei (Hypocrea jecorina).  
OC Zygomycota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN 111  
RP SEQUENCE FROM N. A.  
RC STRAIN=QM9414 / Rut C-30;  
RC MEDLINE=95075308; PubMed=7984103;  
RA Saloheimo A., Henriksa B., Hoffman A.-M., Telesman O., Pectillae M.,  
RT "A novel, small endoglucanase gene, eg15, from Trichoderma reesei  
RT isolated by expression in yeast."  
RL Mol. Microbiol. 13:219-228(1994).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -1- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z33381; CAAB3846.1; -.  
DR PIR; S60143; S60143.  
DR HSSP; P00725; 2CBH.  
DR InterPro: IPR000254; CBD\_fungal.  
DR InterPro: IPR000312; Expan\_endogl.  
DR InterPro: IPR000334; Glyco\_hydro\_45.  
DR Pfam; P500734; CBM\_1; 1.  
DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; FCBH; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
DR PROSITE; PS00842; EXPANSIN\_BG45; 1.  
DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 242  
FT ACT\_SITE 27 27 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 134 134 PROTON DONOR (BY SIMILARITY).  
FT DOMAIN 18 182 CATALYTIC.  
FT DOMAIN 183 205 PRO/SER-RICH (LINKER).  
FT DOMAIN 206 242 CELLULOSE-BINDING (BY SIMILARITY).  
FT CARBOHYD 182 182 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT DISULFID 213 230 BY SIMILARITY.  
FT DISULFID 224 240 BY SIMILARITY.  
SQ SEQUENCE 242 AA; 24411 MW; CC033FC51326CT1D CRC64;  
Query Match 13.3%; Score 221; DB 1; Length 242;  
Best Local Similarity 24.7%; Pred. No. 1.9e-09;  
Matches 78; Conservative 39; Mismatches 101; Indels 98; Gaps 11;

DB 86 APCSSCGTGGAGGSIIMVNLCPNNGAQCPIV--GGTN-----QY- 127  
QY 171 GISSRDQCSFPAPLPGCQRRD-WFQN--ADNPTTPOQVCPALIVANSQ---CKR 223  
DB 128 -----GYSYHEFIMAQNEIRFGDWWVDFEPLACFQAASDMWTCLCVG 170  
QY 224 NDDSPFVFPSPGNGGCTGPTSTAPSGGQTSFGGSGGCSQXMMQCGG1PFGSCCTCV 283  
DB 171 QQETD-P--TPVLGNDTSTPSSPPATSSSPSGGQQT--LYGQCGAGWTEPTTCQ 225  
QY 284 GGTCCCKLNDYSSQCL 299  
DB 226 APGTCKYQNMYSQCL 241  
RESULT 7  
ID GUN1 TRIRE STANDARD; PRT; 513 AA.  
AC P00725;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I) (CBH1) (1,4-beta-cellulohydrolase).  
GN CBH1.  
OS Trichoderma reesei (Hypocrea jecorina), and  
OS Trichoderma koningi.  
OC Zygomycota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453; 55282;  
RN 111  
RP SEQUENCE FROM N. A.  
RC SPECIES=T.reesei; STRAIN=L27;  
RA Shoemaker S., Schweickart V., Ladner M., Gelfand D., Knox S.,  
RT "Molecular cloning of exo-cellulohydrolase I derived from Trichoderma  
RT reesei strain L27."  
RL Biotechnology 1:691-696(1983).  
RN [2]  
RP SEQUENCE FROM N. A.  
RC SPECIES=T.koningii; STRAIN=G-39;  
RX MEDLINE=94100788; PubMed=7764306;  
RA Way T.T., Hsue T.H., Huang L.;  
RT "Molecular cloning and sequence analysis of the cellulohydrolase I  
RT gene from Trichoderma koningii G-39."  
RL Curr. Microbiol. 28:31-39(1994).  
RN [3]  
RP ACTIVE SITE.  
RC SPECIES=T.reesei;  
RA Tome P., Claysens M.;  
RT "Identification of a functionally important carboxyl group in  
RT cellulohydrolase I from Trichoderma reesei."  
RL FBS Lett. 243:239-243(1989).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS) OF 18-452.  
RX MEDLINE=94310436; PubMed=8036495;  
RA Dwyne C., Staehberg U., Reinikainen T., Ruohonen L., Petersson G.,  
RT Knowles J.K.C., Teeri T.T., Jones T.A.;  
RT "The three-dimensional crystal structure of the catalytic core of  
RT cellulohydrolase I from Trichoderma reesei."  
RL Science 265:524-528(1994).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 18-452.  
RC SPECIES=T.reesei; STRAIN=QM9414 / Rut C-30;  
RX MEDLINE=98128795; PubMed=9466911;  
RA Dwyne C., Staehberg U., Teeri T.T., Jones T.A.;  
RT "High-resolution crystal structures reveal how a cellulose chain is  
RT bound in the 50 A long tunnel of cellulohydrolase I from Trichoderma  
RT reesei."  
RL J. Mol. Biol. 275:309-325(1998).  
RN [6]  
RP STRUCTURE BY NMR OF 478-513.

RC SPECIES=I.reesei;  
RX MEDLINE=90057416; PubMed=2554967;  
RA Krasulis P.J., Clore G.M., Nilges M., Jones T.A., Petersson G.,  
RA Krasulis J., Gronenborn A.M.;  
RT "Determination of the three-dimensional solution structure of the C-  
terminal domain of cellobiohydrolase I from Trichoderma reesei. A  
RT study using nuclear magnetic resonance and hybrid distance  
RT geometry-dynamical simulated annealing.";  
RL Biochemistry 28:7241-7257(1989).  
RN [7]  
RP STRUCTURE BY NMR OF 478-513.  
RC SPECIES=I.reesei;  
RX MEDLINE=91194052; PubMed=3041630;  
RA Martinen M.L., Konttinen T., Kerovuo J., Linder M., Annala A.,  
RA Lindberg G., Reinkainen T., Drakenberg T.;  
RT "Three-dimensional structures of three engineered cellulose-binding  
RT domains of cellobiohydrolase I from Trichoderma reesei.";  
RL Protein Sci. 6:294-303(1997).  
CC -I- FUNCTION: The biological conversion of cellulose to glucose  
CC generally requires three types of hydrolytic enzymes: (1)  
CC Exoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
CC Exocellulohydrolases that cut the disaccharide cellobiose from  
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-  
CC glucosidases which hydrolyze the cellobiose and other short cello-  
CC oligosaccharides to glucose.  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
CC in cellulose and cellotetraose, releasing cellobiose from the non-  
CC reducing ends of the chains.  
CC -I- SUBCELLULAR LOCATION: Secreted.  
CC -I- MISCELLANEOUS: T.reesei produces two different  
CC exocellulohydrolases. They are unique in that they can hydrolyze  
CC crystalline cellulose in the absence of endoglucanases.  
CC -I- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
CC hydrolases).  
CC -I- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).  
CC -----  
DR EMBL: X69376; CAA49596.1; -.  
DR PIR: A03302; KUTCI-  
DR PDB: 1CEU; 01-NOV-94.  
DR PDB: 2CEU; 12-MAR-97.  
DR PDB: 3CEU; 12-MAR-97.  
DR PDB: 4CEU; 12-MAR-97.  
DR PDB: 5CEU; 24-DEC-97.  
DR PDB: 6CEU; 24-DEC-97.  
DR PDB: 7CEU; 24-DEC-97.  
DR PDB: 1CBH; 15-JAN-90.  
DR PDB: 2CBH; 15-JAN-90.  
DR PDB: 1AZ6; 08-APR-98.  
DR PDB: 1AZH; 08-APR-98.  
DR PDB: 1AZJ; 29-APR-98.  
DR PDB: 1AZK; 29-APR-98.  
DR PDB: 1DY4; 18-DEC-00.  
DR PDB: 1BGN; 16-MAY-01.  
DR PDB: 8CEU; 25-FEB-98.  
DR InterPro: IPR000254; CBD\_fungal.  
DR InterPro: IPR008985; ConA\_like\_1ec.g1.  
DR InterPro: IPR001722; Glyco\_hydro\_7.  
DR Pfam: PF00734; CBM\_1.1.  
DR Pfam: PF00840; Glyco\_hydro\_7.1.  
DR ProDom: PD001821; CBD\_fungal.1.  
DR ProDom: PD186135; Glyco\_hydro\_7.1.  
DR SMART: SMO0236; E\_CBD.1.  
DR PROSITE: PS00562; CBD\_FUNGAL.1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;  
KW 3D-structure; Pyrrolidone carboxylic acid.

FT SIGNAL	1	17	
FT CHAIN	18	513	EXOGLUCANASE I.
FT DOMAIN	18	453	CATALYTIC.
FT DOMAIN	454	477	LINER
FT DOMAIN	478	513	CELULOSE-BINDING.
FT ACT_SITE	143	143	PROBABLE.
FT ACT_SITE	229	229	NUCLEOPHILE.
FT ACT_SITE	234	234	PROTON DONOR.
FT MOD_RES	18	18	PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD	62	62	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	81	81	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	287	287	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	401	401	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID	21	89	
FT DISULFID	36	42	
FT DISULFID	67	88	
FT DISULFID	78	84	
FT DISULFID	155	414	
FT DISULFID	189	227	
FT DISULFID	193	226	
FT DISULFID	247	273	
FT DISULFID	255	260	
FT DISULFID	278	348	
FT DISULFID	485	502	
FT DISULFID	496	512	
FT STRAND	19	20	
FT STRAND	30	36	
FT TURN	38	39	
FT STRAND	42	51	
FT HELIX	53	55	
FT STRAND	58	60	
FT TURN	61	62	
FT STRAND	66	66	
FT STRAND	68	69	
FT TURN	70	71	
FT STRAND	72	73	
FT TURN	75	77	
FT HELIX	81	87	
FT STRAND	88	90	
FT TURN	95	99	
FT STRAND	101	104	
FT TURN	105	106	
FT STRAND	107	111	
FT STRAND	113	115	
FT STRAND	119	121	
FT STRAND	123	130	
FT TURN	131	132	
FT STRAND	133	134	
FT STRAND	136	137	
FT TURN	140	141	
FT STRAND	142	149	
FT TURN	151	152	
FT TURN	155	156	
FT STRAND	157	164	
FT TURN	168	174	
FT TURN	176	177	
FT HELIX	182	184	
FT TURN	185	185	
FT STRAND	190	190	
FT TURN	191	192	
FT STRAND	199	200	
FT TURN	201	202	
FT STRAND	203	204	
FT TURN	207	208	
FT STRAND	210	211	
FT TURN	215	217	
FT STRAND	219	220	
FT STRAND	223	226	
FT STRAND	229	235	
FT STRAND	240	245	
FT STRAND	247	247	
FT STRAND	253	256	

Query Match 13.2%; Score 219.5; DB 1; Length 513;  
 Best Local Similarity 26.9%; Pred. No. 5.3e-09;  
 Matches 91; Conservative 16; Mismatches 122; Indels 83; Gaps 22;

QY 16 PIVASASGSGSTRYWDCCKPSCAMPKAAVSQPV--YACDANFQRLSDFFVSGGNG 73  
 DB 211 PSSNNANTGIG--GHSCCSEMDMEAN-SISEALTIPHCCTTGGEICF--GGC--G 261  
 QY 74 SAY-----C--ADQTPAVNDNLAYGFATSIAGSSSSWCCAYALFTSGEVA 122  
 DB 262 GTYSIDNRVGGCTDPCDCMNPRLGNTSFTG-----PSSS-----FTLDITK-- 303  
 QY 123 GRTMVYSTIGDGLG----SNQFDIAMPGGVIGINQ-----CS---SQGLPBAQ 166  
 DB 304 -FLTVVTCFETSGAIRRYVYVQKGTFOQPAELGSGYSGNELNDYCAIEAEFGSSPSD 362  
 QY 169 YGSISSRQDCSPFALPKPGCCMRPFQWQ--ADNPFTFOVQCAFIYVARSCKEND 225  
 DB 363 KGGTLQFKKATSGGWLWMSL-WD-DYANMLMDSTYPTMETSTSGAN--RSGCSTSS 418  
 QY 226 -----DSSFP-----VTP-----PSGGN--GGTGTPTSTAPSSQGTSPGGG 261  
 DB 419 GVPAGVESQSPNAKVTFSNKKGPISGTCNPSGPNPGNPGNRTTTRPA--TTTGGSP 475  
 QY 262 GCTSGKMAQGGTGGSGCTTCVSGTTCOKXNDYSOCL 239  
 DB 476 GPTQSHYGGCGGIGYSGPTVCAHGTCVAINPTYSOCL 513

RESULT 8  
 GUN1\_TRIV1 STANDARD; PRT; 513 AA.

AC P19355;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase) (1,4-  
 DE beta-cellobiohydrolase).  
 GN CBH1.  
 OS Trichoderma viride.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Trichoderma.  
 OX NCBI\_TaxID=5547;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91016856; PubMed=2216737;  
 RA Cheng C., Tsukagoshi N., Uda K. S.;  
 RT "Nucleotide sequence of the cellobiohydrolase gene from Trichoderma  
 RT viride";  
 RL Nucleic Acids Res. 18:5559-5559 (1990).

CC -!- FUNCTION: The biological conversion of cellulose to glucose  
 CC generally requires three types of hydrolytic enzymes: (1)  
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
 CC Exocellobiohydrolases that cut the disaccharide cellobiose  
 CC from the nonreducing end of the cellulose polymer chain; (3)  
 CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other  
 CC short cello-oligosaccharides to glucose.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellotetraose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -!- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
 CC hydrolases).  
 CC -!- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC EMBL; X53931; CAA37878.1; -

DR PTR: S11439; S11439.  
 DR HSSP; P00725; ACBL.  
 DR InterPro; IPR000254; CBD\_fungal.  
 DR InterPro; IPR008985; Cons\_like\_1ec\_g1.  
 DR InterPro; IPR001722; Glyco\_hydro\_7.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR -Pfam; PF00840; Glyco\_hydro\_7; 1.  
 DR ProDom; P0001821; CBD\_fungal; 1.  
 DR ProDom; P0186135; Glyco\_hydro\_7; 1.  
 DR SMART; SM00236; ICB1; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 513  
 FT DOMAIN 18 453  
 FT DOMAIN 454 477  
 FT DOMAIN 478 513  
 FT ACT SITE 229 229  
 FT ACT SITE 234 234  
 FT CARBOHD 62 62  
 FT CARBOHD 81 81  
 FT CAROHD 287 287  
 FT DISULFD 485 502  
 FT DISULFD 496 512  
 SQ SEQUENCE 513 AA; 53891 MW; 58F552717C4C8E CRC64;

Query Match 13.0%; Score 215.5; DB 1; Length 513;  
 Best Local Similarity 26.2%; Pred. No. 1e-08;  
 Matches 86; Conservative 16; Mismatches 137; Indels 69; Gaps 16;

QY 16 PIVASASGSGSTRYWDCCKPSCAMPKAAVSQPV--YACDANFQRLSDFFVSGGNG 73  
 DB 211 PSSNNANTGIG--GHSCCSEMDMEAN-SISEALTIPHCCTTGGEICF--GGC--G 266  
 QY 74 SAY--SC--ADQTPAVNDNLAYGFATSIAGSSSSWCCAYALFTSGPVAGKTV 127  
 DB 267 DRVGTCDDPDGDMVPRLGNTSFTG-----PSSS-----FTLDITK--KLTV 307  
 QY 228 YOSTSTGDLG--SNQFDIAMPGGVIGINQ-----SQGLPBAQVGGIS 173  
 DB 308 VTPFTSGAIRRYVYVQKGTFOQPAELGSGYSGNELNDYCAIEAEFGSSPSDGGT 367  
 QY 174 SRQDCSPFALPKPGCCMRPFQWQ--ADNPFTFOVQCAFI--VARSCKEND 226  
 DB 368 QFKKATSGGWLWMSL-WD-DYANMLMDSTYPTMETSTSGAN--RSGCSTSS 425  
 QY 227 SSSP-----VTP-----PSGGN--GGTGTPTSTAPSSQGTSPGGGCTSGKAAQC 271  
 DB 426 SNSPNKAVVYSNKKGPISGTCNPSGPNPGNPGNPGNRTTTRPA--TTTGGSP 485  
 QY 272 GGTGSGCTTCVSGTTCOKXNDYSOCL 299  
 DB 486 GGTGSGPTVCAHGTCVAINPTYSOCL 513

RESULT 9  
 GUN1\_TRIV1 STANDARD; PRT; 459 AA.

AC P07981;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase EC-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 GN CBH1.  
 OS Trichoderma reesei (Hypocrea jecorina).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 OX NCBI\_TaxID=51453;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=VT-D-80133;  
 RX MEDLINE=87106822; PubMed=2948877;



FT HELIX 367 373  
 FT TURN 375 376  
 FT STRAND 378 387  
 FT TURN 389 392  
 SQ SEQUENCE 459 AA; 48208 MW; D235A256P8C8CB9 CRCE4;

Query Match 12.9%; Score 214; DB 1; Length 459;  
 Best Local Similarity 27.0%; Pred. No. 1.2e-08;  
 Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

QY 19 ASASGSGQS---TRYWDCCKSPGKAAVQPYACDA-----NFORLSDENV 66  
 DB 101 ASGVTTSSSLTMYQY---MPSSS-GGYSVSPRLYLDSDGKVMKNGEIS-PDV 154  
 QY 67 Q-----SCGSSGSAVSCADQTP--NAVNDNLAYGFA 95  
 DB 155 DLSALPCGENSLYLTSQMDENGANQVNTAGAVNGGY-CLAQGVQVQW-NQTL----- 207  
 QY 96 ATSIAGSESSKCC-----AC-----YALITFS--GPV 121  
 DB 208 ----NTSHGFCCEMDILBENSANALTPHSCTATACDSAGCGFNPYSGYKSYTGP- 262  
 QY 122 AGKTVWQST-----STGGDLS-----NOFDI--AMPGGVGHENGCSQ 160  
 DB 262 -GDVDSKTFITITQENTDNGSPGNVSTIKKQNGVDLPASQSGDT--ISCTPS- 317  
 QY 161 FGLPAGQYGGISRDQCSFPAPLKPCCQWRD-WFQNAQ--NFTTFQOVQC----- 211  
 DB 318 ----ASAYGGLATMGKA-----LSSGMVIVFSIWNQNSQYMWLDSGNAAGCSSTBEN 366  
 QY 212 PAIYARSGCKNDSSFPVTPPGGNGGNGTPTTA---PSGGT-----SSQG 259  
 DB 367 PSNIIA-----NNPNTHVFSNIRMGDIGS-TTNSIAPPPPPASTFTSTRSSTYS 419  
 QY 260 GSGCTSQKMAQCGGIGPSGCTTCVGGTTCQKLANDYSSQCL 299  
 DB 420 SPSCQTHMGQCGSIGSGCKTCTSGTTCCQYNDYSQCL 459

RESULT 10  
 ID GUN1 TRILO STANDARD; PRT; 463 AA.  
 AC 012714;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Endoglucanase EG-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGI1.  
 OS Trichoderma longibrachiatum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Trichoderma.  
 OX NCBI\_TaxID=5548;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=CECT 2606;  
 RA Perez-Gonzalez J. A.;  
 RL Submitted (NOV-1992) to the EMBL/GenBank/DDBJ databases.  
 CC -I- FUNCTION: The biological conversion of cellulose to glucose  
 CC generally requires three types of hydrolytic enzymes: (1)  
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
 CC Exocellulohydrolases that cut the disaccharide cellobiose  
 CC from the nonreducing end of the cellulose polymer chain; (3)  
 CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other  
 CC short cello-oligosaccharides to glucose.  
 CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -I- STRUCTURAL LOCATION: Secreted.  
 CC -I- SIMILARITY: Belongs to cellulase family C (family 7 of glycoyl  
 CC Hydrolases).  
 CC -I- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announe/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X60652; CAA43059.1; -;  
 DR HSRP; A48375; A48375.  
 DR HSRP; P07981; IBL1.  
 DR InterPro; IPR000254; CBD\_fungal.  
 DR InterPro; IPR008985; ConA\_like\_1ec\_g1.  
 DR InterPro; IPR001722; Glyco\_hydro\_7.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR Pfam; PF00840; Glyco\_hydro\_7; 1.  
 DR ProDom; PD186135; Glyco\_hydro\_7; 1.  
 DR SMART; SM00236; fCBD; 1.  
 DR ProSite; PS00562; CBD\_FUNGAL; 1.  
 KM Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 463  
 FT DOMAIN 23 397  
 FT DOMAIN 402 427  
 FT DOMAIN 428 463  
 FT ACT\_SITE 218 218  
 FT ACT\_SITE 223 223  
 FT DISULFID 435 452  
 FT DISULFID 446 462  
 FT CARBOHYD 78 78  
 FT CARBOHYD 164 164  
 FT CARBOHYD 204 204  
 FT CARBOHYD 208 208  
 FT CARBOHYD 394 394  
 SQ SEQUENCE 463 AA; 48337 MW; B3AC3DPD3ADZBLC CRCE4;

Query Match 12.2%; Score 203; DB 1; Length 463;  
 Best Local Similarity 28.4%; Pred. No. 7.5e-06;  
 Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

QY 68 SGCGSGSAVSCADQTP--NAVNDNL--AYGFAAT-SIAGSESSKCCACVATPFSG 119  
 DB 184 AGAYNSGCV-CDAQCPVQVQW-NQTLNTSGGFCCEMDILBENSANALTPHSCTATAC 241  
 QY 120 PVAG-----KTMVY-----QSTSTGDLGS-----NOFD 143  
 DB 242 DSACGCFNPYSGGYNYPFGDVTWTSKTFITITQENTDNGSPGNVSTIKKQNGVD 301  
 QY 144 I--AMPQGVGIFNGCSSQFGGLPGAQYGGISRDQCSFPAPLKPCCQWRD-WFQNAQ 280  
 DB 302 IPSAKPGSDT--ISSCPB-----ASAYGGLATMGKA-----LSSGMVIVFSIWNQNSQ 347  
 QY 201 -----NPTTFQOVQCPAIVASGCKNDSSFPVTPPGSN--- 239  
 DB 348 YMWMLDSGRAGPCSSTEGNPNIL--ANNPQTHVIVSNIRMGDIGS--TTNSGSGNPP 402  
 QY 240 ---GATGTPSTPSPSGQTPSGGSGCTSQKMAQCGGIGSGCTTCVSGTTCQKLANDYSS 236  
 DB 403 PPPPASTFTSTTRSSSTTS--SSPSCQTHMGQCGSIGSGCKTCTSGTTCCQYNDYSS 460  
 QY 297 QCL 299  
 DB 461 QCL 463  
 RESULT 11  
 ID GUN1 HOMGE STANDARD; PRT; 525 AA.  
 AC P15828;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DB Exoglucanase I precursor (EC 3.2.1.91) [Exocellulohydrolase I] (1,4-  
 DB beta-cellobiohydrolase) (Beta-glucanase; cellobiohydrolase).

GN CBR-1.  
OS Humicola grisea var. thermoides.  
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
OX NCBI\_TaxID=5528;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90175006; PubMed=2308855;  
RA de Oliveira Alzevedo M., Radford A.;  
RT "Sequence of cbh-1 gene of Humicola grisea var. thermoides.";  
RL Nucleic Acids Res. 18:668-668(1990).  
CC -1- FUNCTION: The biological conversion of cellulose to glucose  
CC generally requires three types of hydrolytic enzymes: (1)  
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
CC Exocellulohydrolases that cut the disaccharide cellobiose from  
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-  
CC glucosidases which hydrolyze the cellobiose and other short cello-  
CC oligosaccharides to glucose.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
CC in cellulose and celotetraose, releasing cellobiose from the non-  
CC reducing ends of the chains.  
CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X17258; CA35159.1; -.  
DR PIR; S38794; S38794.  
DR HSRP; P00725; 8CEU.  
DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR008985; Com1\_like\_1ec\_g1.  
DR InterPro; IPR001722; Glyco\_hydro\_7.  
DR Pfam; PF00734; CBM\_1; 1.  
DR Pfam; PF00840; Glyco\_hydro\_7; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR ProDom; PD186135; Glyco\_hydro\_7; 1.  
DR SMART; SMO0236; fCBD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
KM Cellulose degradation: Hydrolyase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 525  
FT DOMAIN 19 467  
FT DOMAIN 468 489  
FT DOMAIN 490 525  
FT ACT\_SITE 231 231  
FT ACT\_SITE 236 236  
FT CARBOHYD 289 289  
FT DISULFID 497 514  
FT DISULFID 508 524  
SQ SEQUENCE 525 AA; 55693 MW; A6884D4C8E81B090 CRC64;  
Query Match 11.2%; Score 185; DB 1; Length 525;  
Best Local Similarity 24.9%; Pred. No. 1.7e-06;  
Matches 97; Conservative 33; Mismatches 136; Indels 124; Gaps 23;  
QY 11 LAAALPLVASASGSGSTRYMDCKKSCAMPK-KAASQPVYACDAPFQLSPF----- 64  
DB 159 LAGALYFVSMADDS--LSRY-----FGKKAQKXGYCYDAGCPDIXFINGRA 206  
QY 65 NVDSGNC-----GGNYS--CADQTPMAYNDMLAYGR--NTSIAGGS--ESSWCAC 111  
DB 207 NIEGWTGSTNDPNAGARYGTCCESEMDIWEAN--NMAAFEPHPTIIQSRCBDSGCGT 265  
QY 112 YALFTPTSG-----FVAGKTVVOST-----STGGDGS--NOF 142  
DB 266 YSNERYAGVCPDGGCDNSYRQGNKTFYKGNKTVDTTKITVYQFLDANGDLGEIKR 325

QY 143 -----DIAMPG--GGVGFNCGSSO---FGGLPG-AQYGGISRDQCSFPAFL 185  
DB 326 YVQDKIIIPSESTTYGEGNSITDWCGRQKARFDDIDPRKRGKQKQKALAGPML 385  
QY 186 KPGCWRFDFQONADNPTF---TEQVQCPAEIVASGCKEN---DDSEFV----- 231  
DB 386 VMS-----WDIASMMMLDSTF-----FVDAAGKPAERAGCAPTTSQVPAVEAEAPN 435  
QY 232 -----FTP-----DSGNGTG-----TPSTAQSQTSPGSGSCTSQKMA 263  
DB 436 SNVFSNRIRGPIGISTVAGLPAGAGNGNNGNPNPTTTSASAPATTTASAPKRWQ 495  
QY 270 QCGGIFSGGCTTCVSGTTCQKNDYYSQCL 299  
DB 496 QCGGIFGTFPTQCEBPYICTKNDYISQCL 525  
RESULT 12  
ID GUXL PHACH STANDARD; PRT; 516 AA.  
AC P13860;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)  
DE (1,4-beta-cellobiohydrolase).  
GN CHN1.  
OS Phanerochaete chrysosporium.  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Aphyllophorales; Corticiaceae; Phanerochaete.  
OX NCBI\_TaxID=5306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 34541 / ME446;  
RX MEDLIN=89232732; PubMed=3246351;  
RA Sims P.F.G., James C., Broda P.;  
RT "The identification, molecular cloning and characterisation of a gene  
RT from Phanerochaete chrysosporium that shows strong homology to the  
RT exo-cellobiohydrolase I gene from Trichoderma reesei.";  
RL Gene 74:411-422(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 34541 / ME446;  
RX MEDLIN=9435641; PubMed=8057846;  
RA Sims P.F.G., Soares-Felipe M.S., Wang Q., Gent M.E., Tempelaar C.,  
RA Broda P.;  
RT "Differential expression of multiple exo-cellobiohydrolase I-like  
RT genes in the lignin-degrading fungus Phanerochaete chrysosporium.";  
RL Mol. Microbiol. 12:209-216(1994).  
CC -1- FUNCTION: The biological conversion of cellulose to glucose  
CC generally requires three types of hydrolytic enzymes: (1)  
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
CC Exocellulohydrolases that cut the disaccharide cellobiose  
CC from the nonreducing end of the cellulose polymer chain; (3)  
CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other  
CC short cello-oligosaccharides to glucose.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
CC in cellulose and celotetraose, releasing cellobiose from the non-  
CC reducing ends of the chains.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M22220; AAB46373.1; -.

DR EMBL: Z22528; CAA80253.1; -  
 DR PIR: J00083; J00083.  
 DR HSSP: P00725; 2CBH.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR008985; ConA\_like\_1ec.g1.  
 DR InterPro: IPR001722; Glyco\_hydro\_7.  
 DR Pfam: PF00734; CBM 1; 1.  
 DR Pfam: PF00840; Glyco\_hydro\_7; 1.  
 DR ProDom: PD01821; CBD\_fungal; 1.  
 DR ProDom: PD186135; Glyco\_hydro\_7; 1.  
 DR SMART: SM00236; FCBP; 1.  
 DR ProSITE: PS00562; CBD\_FUNGAL; 1.  
 XW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 516  
 FT DOMAIN ? 449  
 FT DOMAIN 450 480  
 FT ACT\_SITE 481 516  
 FT ACT\_SITE 225 225  
 FT ACT\_SITE 230 230  
 FT CARBOHYD 208 208  
 FT CARBOHYD 326 326  
 FT CARBOHYD 442 442  
 FT DISULFID 488 505  
 FT DISULFID 499 515  
 FT COMPLECT 27 28  
 FT COMPLECT 30 31  
 FT COMPLECT 31 31  
 SQ SEQUENCE 516 AA; 54857 MW; 1C7C3D38CE1B72 CRC64;  
 Query Match 11.1%; Score 184; DB 1; Length 516;  
 Best Local Similarity 24.7%; Pred. No. 2e-06;  
 Matches 86; Conservative 31; Mismatches 113; Indels 112; Gaps 20;  
 QY 20 SAASGSGSTRYWDCKPCSCAMPK--AAVSQPYACDANFQRLSDFNVSQGNCG-- 73  
 DB 211 SNAAGT---NYGTCTEMDMEANNDAAVTP-HPCT-----NAQTRCSGSDCTR 258  
 QY 74 -----SAVCAQDTFPAVNDLVANGFAISIASGSESSWCACALTFSSGV 121  
 DB 259 DTGLCADGCDNFNSFRMGDT-----FLKGLTVDDSKPFTVTQFTT-NDGTS 306  
 QY 122 AG-----KTVVVGSTGTGDLGSGNDFIAMEG-AGVGIF--NGCSSQ---EGGIP-CAQY 169  
 DB 307 AGTLTEIRLTVQN---GVYONS-SYKIGEDIPVNSITDFFCQCKTARQDNYFAQH 361  
 QY 170 GGSISRDQC-----DSFAPLPKPCQMRFEWQ-----NADNPTFTFGVQC- 211  
 DB 362 GELQVGEALRTGVVALS-WDQYANML-----WLDQVYPTNKDPTSPGVARGTCA 413  
 QY 212 -----PAEIVARSGCKRNDSSFPFTPPSGNGCT-----GPTSTAP 250  
 DB 414 TTSGVPAQIEAQ-----FNAYVFSNIRKGDALNTYGTGVSSSSVSSHSSTJSSSH 467  
 QY 251 GSGQSTPFGGSGCTSCKMAQCGIGFSGCTTCEVETTCQKLDNDYSQC 298  
 DB 468 SSSSTFTGPTGVTTPQMGCGSIGTGSTTCAAPYCAVLAIPYISQC 515  
 RESULT 33  
 ID GTXI PENJA STANDARD; PRT; 537 AA.  
 AC 006886; 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DR Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)  
 DB (1,4-beta-cellobiohydrolase).  
 GN CBH1.  
 OS Penicillium janthinellum (Penicillium vitale).  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eubacterales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
 CX NCBI\_TaxID=5079;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C41;  
 RX MEDLINE=93178976; PubMed=8440481;  
 RA Koch A., Weigel C.T.O., Schult G.;  
 RT "Cloning, sequencing, and heterologous expression of a cellulase-  
 encoding cDNA (cbh1) from Penicillium janthinellum";  
 RL Gene 124:57-65(1993).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 in cellulose and cellobiose, releasing cellobiose from the non-  
 reducing ends of the chains.  
 CC -1- SUBCELLULAR LOCATION: secreted.  
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
 hydrolases).  
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X59054; CAA1780.1; -  
 DR PIR: J00150; J00150.  
 DR HSSP: P00725; 8CEU.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR008985; ConA\_like\_1ec.g1.  
 DR InterPro: IPR001722; Glyco\_hydro\_7.  
 DR Pfam: PF00734; CBM 1; 1.  
 DR Pfam: PF00840; Glyco\_hydro\_7; 1.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR ProDom: PD186135; Glyco\_hydro\_7; 1.  
 DR SMART: SM00236; FCBP; 1.  
 DR ProSITE: PS00562; CBD\_FUNGAL; 1.  
 KV Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 537  
 FT DOMAIN 19 453  
 FT DOMAIN 454 477  
 FT DOMAIN 478 537  
 FT ACT\_SITE 235 235  
 FT ACT\_SITE 240 240  
 FT ACT\_SITE 136 136  
 FT CARBOHYD 414 414  
 FT CARBOHYD 456 456  
 FT DISULFID 509 526  
 FT DISULFID 520 536  
 SQ SEQUENCE 537 AA; 56842 MW; A6B9C6EB73F17FE4 CRC64;  
 Query Match 11.0%; Score 183; DB 1; Length 537;  
 Best Local Similarity 23.4%; Pred. No. 2.5e-06;  
 Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;  
 QY 24 GSGQSTYWDCKPCSCAMPKAAVSQPYACDANFQRLSDFNVSQGNCGSACADQTP 83  
 DB 268 GTYSTDRVGTQDPD-----CQDNPRKGVNFX-----GPGTIDTFRSP 308  
 QY 84 WAV-----NDNLAVFAA-----TSLAGS-ESSWCACALTF 116  
 DB 309 FTVVTCFELTMDGTSTSLSEIKRFYVQSGKVIQNPSTIVGSGNSITDSWCA----- 362  
 QY 117 TSGPAAKIMVVGSTGTGDLGSGNDFIAMEG-AGVGIF--NGCSSQ---EGGIP-CAQY 169  
 DB 363 -----OKSAPFD-TWESKKGNAQAGLADSVTLVMSLMDHSDMLML 407  
 QY 158 -----SFGQGLPQAQYGGISRRQCDSPAPLPKPCQMRFEWQ-----NADNPTFTFGVQC 212  
 DB 408 DSTYVTAISITTPGK-----KGTCDISRRP-----NTVESTY-----P 441  
 QY 213 AEIVARSGCKRNDSSFPFTPPSGGNGGIGTGTSTAPRSQC-----TSRGGSS 261  
 DB 442 NAYVYVSNIRKGPLNS--TFTGGTSSSSITTTTSKSTSTSSSKXTTIVTTTSSSS 499



DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 320 POTENTIAL.  
FT DOMAIN 30 261 CELLULOSE-GROWTH-SPECIFIC PROTEIN.  
FT DOMAIN 262 285 CATALYTIC (POTENTIAL).  
FT DOMAIN 286 320 LINKER (POTENTIAL).  
FT DISULFID 232 309 CELLULOSE-BINDING (POTENTIAL).  
FT DISULFID 303 319 BY SIMILARITY.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 320 AA; 33754 MW; 60E2C8C80895CA2B CRC64;

Query Match 10.4%; Score 172; DB 1; Length 320;  
Best Local Similarity 24.8%; Pred. No. 9.2e-06;  
Matches 85; Conservative 36; Mismatches 132; Indels 90; Gaps 15;

QY 10 TLAALPLVNSAASGGGOSTRY-WDCC-----KPSGAMPKAAVSPVACDANFQRLSD 63  
DB 15 TFLALGLFAKVGQAHGVIGYSMDGTWYBGMPYNTPVGCTSIERPW----ATFDPIMD 70  
QY 64 FNVQS-GGNGGSAYSQADQTPMAVNDLNGFPAISIAGSSSSWCCACTALTFTISGPVA 122  
DB 71 ATASTVGCNN-----DENP-GPQULTATVAAGTALTAYNQVMPHPYGPMTTYLGKCP 122  
QY 123 GKTWVVGSTST-----GSDLG-----SNQPDIAIPGGVGIFNGCS 158  
DB 123 GSSCGVNTNSLKMFKIDENGLSGTVGKGVWGSQKIDQNNSWTTIIP----- 171  
QY 159 SQFGGLPGAQYGGISRDQDSFPALPKPGCQMRFDWFQNAQNDPTFTFQVQCP--AEIV 216  
DB 172 ---STVPSGAYMIRFETIALHSLPAQIYPECA-----QLITGGGNRAPISSSEIV 218  
QY 217 ARSGGKRD-----DSSFPVTPP---SGNGG--TGTPTSTAPSGGQTS 256  
DB 219 SFPGGYSNSDPGLTVNLVYQEMMDFTYIVGPPPLYSGGNGSSPTTPTPTTPTLTSPP 278  
QY 257 PGSGSGCTSKMAQCGAIGFSGCTTGVSGTTCKLNDYYSCCL 299  
DB 279 PTSTPG-TIPYGGCGSIGMTGTGTCVAPYQCKVINDYISQCL 320

Search completed: July 7, 2004, 08:53:23  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CV protein - protein search, using sw model

Run on: July 7, 2004, 08:47:55 ; Search time 45 Seconds

(without alignments)  
2096.443 Million cell updates/sec

Title: US-10-007-521-12

Sequence: 1 MRSPTVTRTAAALPIVAS.....TTTCUSGTCCKANDYISQL 299

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_Archea:\*  
2: SP\_Bacteria:\*  
3: SP\_Fungi:\*  
4: SP\_Human:\*  
5: SP\_Invertebrate:\*  
6: SP\_Mammal:\*  
7: SP\_rhbc:\*  
8: SP\_Organelle:\*  
9: SP\_Phage:\*  
10: SP\_Plant:\*  
11: SP\_Rodent:\*  
12: SP\_Virus:\*  
13: SP\_Vertebrate:\*  
14: SP\_Unclassified:\*  
15: SP\_Virus:\*  
16: SP\_Bacteriophage:\*  
17: SP\_Archea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1272	76.7	293	097201	097201 neurospora
2	1170	70.5	305	093782	093782 humicola gr
3	1039	62.6	235	0800X8	093088 meliocarpa
4	757.5	45.7	366	3	080110 rhizopus or
5	756.5	45.6	338	3	080112 rhizopus or
6	745	44.9	360	3	080111 rhizopus or
7	724	43.6	219	14	090192 unclassified
8	724	43.6	220	14	090194 unclassified
9	709	42.7	219	14	090191 unclassified
10	706.5	42.6	220	14	090193 unclassified
11	705	42.5	217	14	090195 unclassified
12	703	42.4	219	14	090196 unclassified
13	700	42.2	219	14	090197 unclassified
14	698.5	42.1	227	3	093783 humicola gr
15	698	42.1	219	14	090190 unclassified
16	697.5	42.0	221	14	090194 unclassified

17	697	42.0	219	14	090198 unclassified
18	697	42.0	219	14	090199 unclassified
19	693	41.8	220	14	090196 unclassified
20	689.5	41.6	220	14	090193 unclassified
21	680.5	41.0	219	14	090195 unclassified
22	632	38.1	219	5	070278 unclassified
23	631	38.0	410	3	090196 unclassified
24	625	37.7	219	5	070277 unclassified
25	617.5	37.2	220	5	070279 unclassified
26	613.5	37.0	220	5	070280 unclassified
27	585.5	35.3	271	3	090193 unclassified
28	543.5	32.8	237	5	080195 unclassified
29	543.5	32.8	242	5	090191 unclassified
30	360.5	21.7	112	14	090199 unclassified
31	348.5	21.0	112	14	090196 unclassified
32	347.5	20.9	112	14	090198 unclassified
33	338.5	20.4	112	14	090192 unclassified
34	338.5	20.4	112	14	090191 unclassified
35	330	19.9	111	14	090197 unclassified
36	221	13.3	247	3	070270 unclassified
37	215.5	13.0	459	3	070273 unclassified
38	215	13.0	514	3	093832 unclassified
39	200	12.1	505	3	090193 unclassified
40	191	11.5	525	3	012621 unclassified
41	188.5	11.4	302	3	090194 unclassified
42	188	11.3	293	3	080198 unclassified
43	187.5	11.3	316	3	070299 unclassified
44	183	11.0	523	3	090193 unclassified
45	181.5	10.9	529	3	080274 unclassified

#### ALIGNMENTS

RESULT 1	097201	PRELIMINARY:	PRT;	293 AA.
ID	097201	01-JUN-2003 (TEMBLrel. 24, Created)		
AC	097201	01-JUN-2003 (TEMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TEMBLrel. 25, Last annotation update)			
DE	Hypothetical protein B19A17.010.			
GN	B19A17.010.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
CC	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_TaxID=5141;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Schulte U., Aign V., Hobeisel U., Brandt P., Fartmann B., Holland R.,			
RA	Nyakatura G., Mewes H.W., Mannhaupt G.;			
RU	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	German Neurospora genome project;			
RU	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BX284763; CAD70529.1;			
DR	GO; GO:0005576; C:extracellular; IEA.			
DR	GO; GO:0008810; F:cellulase activity; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR009009; Barwin-like.			
DR	InterPro; IPR000254; CBD_fungal.			
DR	InterPro; IPR000354; Glyco_hydro_45.			
DR	Pfam; PF00734; CBM_1.			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	ProDom; PD001821; CBD_fungal; 1.			
DR	SMART; SMO0236; fCBD; 1.			
DR	PROSITE; PS00552; CBD_FUNGAL; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.			
KW	Hypothetical protein.			
SO	SEQUENCE 293 AA; 30261 MW; 27096407349877A CRC64;			
Query Match	76.7%; Score 1272; DB 3; Length 293;			

Best Local Similarity 74.8%; Pred. No. 1.5e-93;  
Matches 223; Conservative 30; Mismatches 39; Indels 6; Gaps 2;

QY 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 60  
DB 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 60  
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 120  
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 120  
QY 121 VAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 180  
DB 121 VAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 180  
QY 181 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGG 240  
DB 181 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGG 240  
QY 241 GTGTFSTADAGSGQTSPPGSGGCTSGKMAQCCGTFSGCTTCTVSTTCQKANDYSSQ 298  
DB 241 NPSTPT-TPSS-----GGSGCTADKTAQOGSGSGCTNCSGCTCKTINDYHOC 292

## RESULT 2

ID 093782 PRELIMINARY; PRT; 305 AA.  
AC 093782;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Endoglucanase.  
GN Egl3.  
OS Humicola grisea var. thermoides.  
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
OX NCBI\_Taxid=5528;  
RN [1]  
RS SEQUENCE FROM N.A.  
RC STRAIN=IF09854;  
RX MEDLINE=39144540; PubMed=990729;  
RA Takashima S., Iikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;  
RT "Comparison of gene structures and enzymatic properties between two  
RT endoglucanases from Humicola grisea.";  
RL J. Biotechnol. 67:85-97(1999).  
DR EMBL; AB001107; BA074956.1; -.  
DR HSP; P4316; ZENG.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008810; F:cellulase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR009009; Barwin\_like.  
DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR000334; Glyco\_hydro\_45.  
DR Pfam; PF00734; CBM\_1; 1.  
DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; ECRD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
SQ SSQDNCE 305 AA; 32174 MW; 28C979DEDDC771D CRC64;

Query Match 70.5%; Score 1170; DB 3; Length 305;  
Best Local Similarity 67.8%; Pred. No. 2.1e-85;  
Matches 208; Conservative 44; Mismatches 45; Indels 10; Gaps 4;

QY 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 60  
DB 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 60  
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 119  
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 118  
QY 120 PVAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 179

DB 119 PVAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 178  
QY 180 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGG 238  
DB 179 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPSGG 238  
QY 239 NGSTGTFSTADAGSGQTS-----PGSGGCTSGKMAQCCGTFSGCTTCTVSTTCQKXN 292  
DB 239 NGSTGTFSTADAGSGQTS-----PGSGGCTSGKMAQCCGTFSGCTTCTVSTTCQKXN 292  
QY 293 DYSYQCL 299  
DB 299 DYSYQCL 305

## RESULT 3

ID 0810K8 PRELIMINARY; PRT; 235 AA.  
AC 0810K8;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Cellulase precursor (EC 3.2.1.4).  
GN CEL45A.  
OS Melanocarpus albomyces.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariales incertae sedis;  
OC Melanocarpus.  
OX NCBI\_Taxid=204285;  
RN [1]  
RS SEQUENCE FROM N.A.  
RA Haakana H., Miettinen-Oinonen A., Joutsjoki V., Mantyla A.,  
RA Suominen P., Vehmänpää J.;  
RT "Cellulase Genes from Melanocarpus albomyces: Cloning and High-level  
RT Expression in Trichoderma reesei.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ515703; CAD56655.1; -.  
DR GO; GO:0008810; F:cellulase activity; IEA.  
DR GO; GO:0016798; F:hydrolyase activity; acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR009009; Barwin\_like.  
DR InterPro; IPR000334; Glyco\_hydro\_45.  
DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
KM Signal; Hydrolyase; Glycosidase.  
FT SIGNAL 21  
FT CHAIN 22 235  
SQ SEQUENCE 235 AA; 25001 MW; BF9850B8366CD76 CRC64;

Query Match 62.6%; Score 1039; DB 3; Length 235;  
Best Local Similarity 76.3%; Pred. No. 4.2e-75;  
Matches 180; Conservative 27; Mismatches 27; Indels 2; Gaps 1;

QY 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 60  
DB 1 MRSTPVARTTAAALPLVSAASGSGSTRYWDCKPSGKAPKAAVSQPVYACDANFOR 58  
QY 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 120  
DB 61 LSDENVVSGGNGSAYSCADOTPMVANDLAFGAATSIAGSESSWCCACATLTFTSGP 118  
QY 121 VAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 180  
DB 121 VAGKTIVVOSTSTGGDLGSNOFDIAMPGGGAGIEMGSSQFGGLPAQVAGISSRDQCS 178  
QY 181 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPS 236  
DB 181 FPAPLKGCGMRFDWFOANADNPTFTFOVQCPAEIVASGCKRNDSSFPVTPPS 234

## RESULT 4

QY 080110

ID	080110	PRELIMINARY	PRT	366 AA.
AC	080110			
DT	01-MAR-2003	(TREMblrel. 23, Created)		
DT	01-MAR-2003	(TREMblrel. 23, Last sequence update)		
DT	01-OCT-2003	(TREMblrel. 25, Last annotation update)		
DE	Endo-glucanase RCE3.			
CN	RCE3.			
CC	Rhizopus oryzae (Rhizopus delemar).			
CC	Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;			
CC	Rhizopus.			
OX	NCBI_TaxID=64495;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PE2M BP-6899;			
RA	Marashina K.;			
RT	Cloning of new endo-glucanases rce2 and rce3 from Rhizopus oryzae.*;			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB056668; BACS388.1; -			
DR	GO; GO:0005576; C:extracellular; IEA.			
DR	GO; GO:0008810; P:cellulase activity; IEA.			
DR	GO; GO:0008975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR009009; Barvin-like.			
DR	InterPro; IPR000254; CBD_fungal.			
DR	InterPro; IPR000334; Glyco_hydro.45.			
DR	Pfam; PF00734; CBM_1; 2.			
DR	Pfam; PF02015; Glyco_hydro.45; 1.			
DR	ProDom; PD001821; CBD_fungal; 2.			
DR	SMART; SMO0236; fCBD; 2.			
DR	PROSITE; PS00562; CBD_FUNGAL; 2.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.			
DR	SEQUENCE			
DR	366 AA; 3603 FM; 8A22D01BF636E374 CXC64;			

Query Match	45.7%;	Score 757.5;	DB 3;	Length 366;
Best Local Similarity	62.3%;	Pred. No. 1.9e-52;		
Matches :34;	Conservative 32;	Mismatches 40;	Indels 9;	Gaps 4

QY	17	LYAAABAGSGQSTRHDDCCPSSAMRGKALAYQVYACDAN--EQRLSDPEYVSGCNGGSA	75
DB	152	IVSGAGSAGNGVITTRWDCCKASCSMPGRANVSPVASCNKACVCLNDSNVSGCNGGNS	211
QY	76	YSCADQTPAAVNDNLAYGFAATSIAGSSBSMCACATLFTFSGPVGAKTMVQSTGTG	135
DB	212	YMCQKDPAAVNDNLAYGFMAALISGGSESRCCSGELFTFSTSVAGKGVVQVLTNGR	271
QY	136	DIUGSN--QFDIAMPGGSGVGIENGGSSQFGSLP---GAQYGGIISRQDQDSFPAFLKPG	188
DB	272	DIUGSTGAHFELQMPGGGVGIENGGCKQMGCA--PNDGWSRYGGISASDCSSLPALQAG	330
QY	189	COMPEFQMANDPPTFTQQYQCEPAIVAVSGCR	223
DB	332	CMKREHMFKNADPSMTIKVETGCKEKLTKNGCSR	365

RESULT 5			
CEJ112			
ID	Q6J112	PRELIMINARY;	PRT; 338 AA.
AC	Q6J112;		
DT	01-MAR-2003 (TREMBL1, 23, Created)		
DT	01-MAR-2003 (TREMBL1, 23, last sequence update)		
DT	01-OCT-2003 (TREMBL1, 25, last annotation update)		
DE	Endo-beta-1,4-D-glucanase.		
GN	RCE1.		
OS	Rhizopus oryzae (Rhizopus delamar) .		
OC	Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;		
OC	Rhizopus.		
CX	NCBI_TaxID=64495;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FERM BR-6889;		
RA	Murasaka X;		
RT	*Cloning of new endo-beta-1,4-D-glucanase, RCE1 from Rhizopus oryzae		
RL	CP96001";		
RT	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		

Query Match	Similarity	Score	DB 3:	Length
Best local	63.1%	Pred. No. 2, IE-52:		338;
Matches 135;	Conservative 30;	Mismatches 40;	Indels 9;	Gaps 4;

Query Match	45.64;	Score 756.5;	DB 3;	Length 338;
Best Local Similarity	63.14;	Pred. No. 2,1e-52;		
Matches 135;	Conservative 30;	Mismatches 40;	Indels 9;	Gaps 4
QY	18	VASASGSSGSTRYWDCCCKSCAMPGRKASVQPVACDN-FQRLSPFNQSCNGSSAY	76	
Db	125	VSGAGAGNVITRRYWDCCCKSCAMPGRKXNVSPIKSCNDGVTLASISNAQSCNGGNSY	184	
QY	77	SCDDQPMAYNDNLNAYGEAATSLAGSSSSSCACVALFTTSGPAAGTKWVYSTSGAD	136	
Db	185	MCNDQPMAYNDNLNAYGEAATAISGGSSRKSSCCBELFTTSTVAKGKXNVYNTTSGD	244	
QY	137	LGSN---QEDLAMPGGGVIYFNCGSSQFGSLP---GAQYGSISRRQCSPPALPKRC	185	
Db	245	LGSTGAHEDLMPGGGVIYFNCGSSQNGA-PNDGMSYRGYISBASDPSLPSALOGC	303	
QY	190	QMRFDPMQANPCFTFOOYQCPAEIVARSQCR	223	
Db	304	KMRNFMKADNSKTYKEVTCPEKELTKATQCS	337	

```

RESULT 6
SeqJL1
ID SeqJL1 PRELIMINARY; PRT. 360 AA.
AC SeqJL1;
DT 01-MAR-2003 (TrEMBLrel_23, Created)
DT 01-MAR-2003 (TrEMBLrel_23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel_25, Last annotation update)
DE Endo-glucanase RCE2.
OS RCE2.
SN Rhizopus oryzae (Rhizopus delemar).
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
CX NCBI_TextID=64495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FERM BP-6889;
RA Mirashima K.;
RT "Cloning of new endo-glucanases rce2 and rce3 from Rhizopus oryzae.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB056667; BKCS3987.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR009009; Barwin like.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR000334; Glyco_hydro_45.
DR Pfam; PF00734; CEM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR Prodom; PD001821; CBD_fungal_1.
DR SMART; SM00236; FCBP_1.
DR PROSITE; PS00563; CBD_FUNGAL_1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 360 AA; 37427 MW; F4Z63626C8CD64 CRC64;
Query Match 44.9%; Score 745; DB 3; Length 360;
Best Local Similarity 60.4%; Pred. No. 1,8e-51;
Matches 128; Conservative 36; Mismatches 42; Indels 6; Gaps 2

```

```

QY 18 VASAAGSGGSTRYWDCKPSCAMPKAAVOPVYACDANFQRLSDPNVQSGNGSAYS 77
DB 148 VSGASGMEETRYWDCKPSGKADVTSPVGSCKRQKTLADNNTQNGCVGSSSYT 207
QY 78 CADQTPWAVNDNLAYGPAATSIAGSSSSMCACALFTFTSGPVAKTWTWOSTGGDL 137
DB 208 CNDVQPVWYSDLAAGFAAASISGSSSEATWCCACELTFTSTAVKAKMWTQVNTGSDL 267
QY 138 GSN--OFTDAMPGGGAGVIFNGCSSQFGGLP---GAOYGGISRRDQCSFPAPLPGCCW 191
DB 268 GSNNGAHEFLDMPGGGAGVIFNGCATQMGAPDTGMCARYGVSSASDSCSLPSALQACCKK 327
QY 192 RFPMPKADNPFTFTQOVCRAEIVARSCKR 223
DB 328 RFGWPKAADNPFTMTKYQVTCRKALTKSGSCR 359

RESULT 7
QY 09JH92 PRELIMINARY; PRT; 219 AA.
AC 09JH92;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohtsuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045171; BAA98041.1; -
DR HSSP; P43316; 2ENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000334; Barwin_Like.
DR InterPro; IPR000334; Glyco_Hydro_45.
DR Pfam; PF02015; Glyco_Hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 219 AA; 23001 MW; 5F2EB81A6BE926CE CRC64;

Query Match 43.6%; Score 724; DB 14; Length 219;
Best Local Similarity 60.3%; Pred. No. 4,9e-50;
Matches 132; Conservative 26; Mismatches 53; Indels 8; Gaps 4;

QY 11 LAALPLVASAAGSGGSTRYWDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68
DB 2 LVFTLALISVPGDSGRTRTYWDCKPSCAMPKAAVOPVYDQSGGTRVANDTVKS 61
QY 69 GNGGSAVSCADQTPWAVNDNLAYGPAATSIAGSSSSMCACALFTFTSGPVAKTWTW 128
DB 62 ACDSGEGEYMCQAPNPAVNDVAVGFAAACCAG--ESAGACNCELTFTSGPVAKTWTW 120
QY 139 GSTSTGGDLGSGNDPDIAMPGGGAGVIFNGCSSQFGGLP---GAOYGGISRRDQCSFPAP 184
DB 121 QVNTNGGDLGSGNDPDIAMPGGGAGVIFNGCATQMGAPDTGMCARYGVSSASDSCSLPSAL 179
QY 185 LKPGCCMRFPDMPKADNPFTFTQOVCRAEIVARSCKR 223
DB 180 LQAGCCMRFPDMPKADNPFTFTQOVCRAEIVARSCKR 218

RESULT 8
QY 09JH84 PRELIMINARY; PRT; 220 AA.
AC 09JH84;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

```

```

DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohtsuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045179; BAA98049.1; -
DR HSSP; P43316; 2ENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000334; Barwin_Like.
DR InterPro; IPR000334; Glyco_Hydro_45.
DR Pfam; PF02015; Glyco_Hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 220 AA; 23274 MW; 26AF5357512EA061 CRC64;

Query Match 43.6%; Score 724; DB 14; Length 220;
Best Local Similarity 59.8%; Pred. No. 4,9e-50;
Matches 131; Conservative 30; Mismatches 50; Indels 8; Gaps 4;

QY 11 LAALPLVASAAGSGGSTRYWDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68
DB 3 LKPSLCLTSWVGDSGRTRTYWDCKPSCAMPKAAVOPVYDQSGGTRVANDTVKS 62
QY 69 GNGGSAVSCADQTPWAVNDNLAYGPAATSIAGSSSSMCACALFTFTSGPVAKTWTW 128
DB 63 SCDDGDXGMCYDQAPNPAVNDVAVGFAAACCAG--ETGACNCELTFTSGPVAKTWTW 121
QY 129 GSTSTGGDLGSGNDPDIAMPGGGAGVIFNGCSSQFGGLP---GAOYGGISRRDQCSFPAP 184
DB 122 QVNTNGGDLGSGNDPDIAMPGGGAGVIFNGCATQMGAPDTGMCARYGVSSASDSCSLPSG 180
QY 185 LKPGCCMRFPDMPKADNPFTFTQOVCRAEIVARSCKR 223
DB 181 LQAGCCMRFPDMPKADNPFTFTQOVCRAEIVARSCKR 219

RESULT 9
QY 09JH91 PRELIMINARY; PRT; 219 AA.
AC 09JH91;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohtsuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045172; BAA98042.1; -
DR HSSP; P43316; 2ENG.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000334; Barwin_Like.
DR InterPro; IPR000334; Glyco_Hydro_45.
DR Pfam; PF02015; Glyco_Hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 219 AA; 23126 MW; A712B83E3CAB041C CRC64;

Query Match 42.7%; Score 709; DB 14; Length 219;
Best Local Similarity 59.4%; Pred. No. 7,7e-49;
Matches 130; Conservative 28; Mismatches 53; Indels 8; Gaps 4;

QY 11 LAALPLVASAAGSGGSTRYWDCKPSCAMPKAAVOPVYAC--DANFQRLSDPNVQSG 68

```



Db 6 LTVGLSLAESKTRRYWDCCCKSGCEKXANVDKPIDTCANXGTRRVAASNDTVKSGCDG 65  
 QY 73 GSAYSCADQTPAIVNDLAVGFAATSIAGSSSSSWCCACVALLFTTSGPAAKTMVQSTS 132  
 Db 66 GEGYWCYDQTPWVSVDSTYGFFAAACCCGG-BSGACGCGVDLFTTSGPAAKTMVQITN 124  
 QY 133 TGGDLGNSQFDLAMPGGGAGVGFNGCSSPFGGLP-----GAQYGGISRPQCCSPFAPLKG 188  
 Db 125 TGGDLGNSQFDLAMPGGGAGVGFNGCTAQ-SGAPSDGMSRYGVSSRSRSCQLPSGLQAG 183  
 QY 189 CQWRFDFQNDPNDPFTTPOVQCPAETVARSQCR 223  
 Db 184 CQWRFDFQNDPNDPINSVRCPAETIAKTKCR 218

RESULT 13  
 Q9JH87 PRELIMINARY; PRT; 219 AA.  
 AC Q9JH87  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Family 45 cellulase homologue.  
 OS unclassified eukaryotes.  
 OC Eukaryota.  
 NCBI\_TaxID=42452;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ohtsuka K., Ohkuma M., Moriya S., Kudo T.:  
 RT "Diverse genes of family 45 cellulase homologues of the symbiotic  
 KT protists in the hindgut of termite Reticulitermes speratus";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB045176; BAA98046.1; -.  
 DR HSSP: P43316; ZENG.  
 DR GO: GO:0008810; F:cellulase activity; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR009009; Barwin like.  
 DR InterPro: IPR00334; Glyco\_hydro\_45.  
 DR Pfam: PF02015; Glyco\_hydro\_45; 1.  
 DR PROSITE: PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
 SQ SEQUENCE 219 AA; 23134 MW; 4BDFEFC9ACCT72D CRC64;

Query Match 42.2%; Score 730; DB 14; Length 219;  
 Best Local Similarity 57.7%; Pred. No. 4e-48;  
 Matches 124; Conservative 34; Mismatches 49; Indels 8; Gaps 4;

QY 15 LPLVASAAGSGGSTRYWDCCCKSPCAKGAASQPYVAC--DANFQLSPNVQSGNG 72  
 Db 6 LPLVSLADSGRTTRVYDCCCKSGCEKXANVDKPIDTCANXGTRRVAASNDTVKSGCDG 65  
 QY 73 GSAYSCADQTPAIVNDLAVGFAATSIAGSSSSSWCCACVALLFTTSGPAAKTMVQSTS 132  
 Db 66 GEGYWCYDQTPWVSVDSTYGFFAAACCCGG-BSGACGCGVDLFTTSGPAAKTMVQITN 124  
 QY 133 TGGDLGNSQFDLAMPGGGAGVGFNGCSSPFGGLP-----GAQYGGISRPQCCSPFAPLKG 188  
 Db 125 TGGDLGNSQFDLAMPGGGAGVGFNGCTAQ-SGAPSDGMSRYGVSSRSRSCQLPSGLQAG 183  
 QY 189 CQWRFDFQNDPNDPFTTPOVQCPAETVARSQCR 223  
 Db 184 CQWRFDFQNDPNDPINSVRCPAETIAKTKCR 218

RESULT 14  
 Q93783 PRELIMINARY; PRT; 227 AA.  
 AC Q93783  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Endoglucanase.  
 GN EGL4.  
 GN Humicola grisea var. thermoides.

OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
 NCBI\_TaxID=5528;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IF09854;  
 RA MEDLINE=9144540; PubMed=9990729;  
 RA Takahina S., Iikura H., Nakamura A., Hidaka M., Maeki H., Uozumi T.:  
 RT "Comparison of gene structures and enzymatic properties between two  
 RT endoglucanases from Humicola grisea";  
 RL U. Biotechnol. 67:85-97(1999).  
 DR EMBL: AB00108; BAA74957.1; -.  
 DR HSSP: P43316; ZENG.  
 DR GO: GO:0008810; F:cellulase activity; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR009009; Barwin like.  
 DR InterPro: IPR00334; Glyco\_hydro\_45.  
 DR Pfam: PF02015; Glyco\_hydro\_45; 1.  
 DR PROSITE: PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
 SQ SEQUENCE 227 AA; 24240 MW; 873553B76F5C39E4 CRC64;

Query Match 42.1%; Score 698.5; DB 3; Length 227;  
 Best Local Similarity 52.7%; Pred. No. 5.5e-48;  
 Matches 137; Conservative 26; Mismatches 46; Indels 51; Gaps 8;

QY 9 TPLAALPLVASAAGSGGSTRYWDCCCKSPCAKGAASQPYVACDANFQLSD-FVQ 67  
 Db 6 TPLTLPLALAAQSSGRTTRVYDCCCKSPCAKGAAPV-PRTCRDMNPLFQAGNTR 63  
 QY 68 GSCN-GSAYSCADQTPAIVNDLAVGFAATSIAGSSSSWCCACVALLFTTSGPAAKTM 126  
 Db 64 SCDDAGGAATNCSDSPMAVSDILAYGMAVNLASRQWCCACVLLFTTSGPAAKTM 123  
 QY 127 VVQSTSTGDLGNSQFDLAMPGGGAGVGFNGCSSPFGGLP-----GAQYGGISRPQCCSPFAPLKG 183  
 Db 124 TVQASNTGDLGNSQFDLAMPGGGAGVGFNGCTAQ-SGAPSDGMSRYGVSSRSRSCQLPSGLQAG 183  
 QY 184 PLKPCQWRDFQNDPNDPFTTPOVQCPAETVARSQCRDSSFPVTPPSGGNGTG 243  
 Db 184 KXKPCQWRDFQNDPNDPFTTPOVQCPAETVARSQCRDSSFPVTPPSGGNGTG 204  
 QY 244 TPTSTAGSGGSTRYWDCCCKSPCAKGAASQPYVACDANFQLSD-FVQ 67  
 Db 205 -LSLPLPQVQTM--GRSC 219

RESULT 15  
 Q9JH90 PRELIMINARY; PRT; 219 AA.  
 AC Q9JH90  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Family 45 cellulase homologue.  
 OS unclassified eukaryotes.  
 OC Eukaryota.  
 NCBI\_TaxID=42452;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ohtsuka K., Ohkuma M., Moriya S., Kudo T.:  
 RT "Diverse genes of family 45 cellulase homologues of the symbiotic  
 KT protists in the hindgut of termite Reticulitermes speratus";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB045173; BAA98043.1; -.  
 DR HSSP: P43316; ZENG.  
 DR GO: GO:0008810; F:cellulase activity; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR009009; Barwin like.  
 DR InterPro: IPR00334; Glyco\_hydro\_45.  
 DR Pfam: PF02015; Glyco\_hydro\_45; 1.  
 DR PROSITE: PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
 SQ SEQUENCE 219 AA; 23037 MW; 372E01641530A5A CRC64;

Query Match 42.1%; Score 698; DB 14; Length 219;

Best Local Similarity 56.7%; Pred. No. 5.8e-48;  
Matches 122; Conservative 36; Mismatches 49; Indels 8; Gaps 4;

Matches 122; Conservative 36; Mismatches 49; Indels 8; Gaps 4.

QY 15 LPLVASAGSGSTRYWDCCKPSCAMPGKAASQPYAC--DANFRLSDFNVSQGCNG 72

Db 6 LTFIGLSLADSGKTR YWDCCKGSGWEKKANVDKPICTCAKDGTRVASNDTVKSGCDG 65

73 GSAYS CADQTPWAVNDNLAYGFAATSIAGSESSWCACYALTFTSGPVAGKTMVQSTS 132

66 GTGFMCYDQTPWQVSDSLSYGFAAAACCGG-ESGACCGCYELTFTSGPVNGKMIIVQITN 124

QY 133 TGGDLGSNQFDIAMPGGCVGFNGCSSQFGGL?---GAQYGGISSRDQCCDSFPAPLKPG 188

Db 125 TGGDLGSNQFDLAI PGGVG IYNGCTAQ-SGA?SDGWSR YGGVSSRS ECSQLPSGLQAG 183

QY 189 CQWREDFQNDNPTFTFOQVQCPAEIVARSGCKR 223

Db 184 CQREDFQNDPSINFSNVKCPSEIIAKTNCNR 218

Search completed: July 7, 2004, 08:54:20  
Job time : 46 secs

Job time : 46 secs